An integer-valued bilinear type model

Alain Latour, Lionel Truquet

▶ **To cite this version:**

| Alain Latour, Lionel Truquet. An integer-valued bilinear type model. 2009. <hal-00373409>

HAL Id: hal-00373409
https://hal.archives-ouvertes.fr/hal-00373409

Submitted on 5 Apr 2009

AN INTEGER-VALUED BILINEAR TYPE MODEL

ALAIN LATOUR,* *Université Pierre Mendès-France – LJK-MS*[3]

LIONEL TRUQUET,** *Université de Paris 1 – CREST*

Abstract

A integer-valued bilinear type model is proposed. It can take positive as well as negative values. The existence of the process is established in $\mathbb{L}^m$. In fact, this process is the unique causal solution to an equation that is similar to a classical bilinear type model equation. For the estimation of the parameters, we suggest a quasi-maximum likelihood approach. The estimator is strongly consistent and asymptotically normal.

Keywords: Times series; Bilinear Processes; Stable Processes; Point estimation; Asymptotic properties of estimators; Asymptotically normal estimators;

2000 Mathematics Subject Classification: Primary 62M10; 60G52
Secondary 62F10; 60F12; 60F05;60F15

1. Introduction

As pointed in [19], integer-valued times series are common in practice. In epidemiology, we often consider the number of cases of a given disease over a 28-day period. In this context, the data are collected to make sure that the population is not threatened by an epidemic. As well as in intensive care monitoring, where vital parameters have to be analyzed online, good modeling is required. As soon as three consecutive values seem to be too high, governmental actions are planed to avoid the widespread of the disease, since there may be serious consequences for the population otherwise. See [12] where regression methods are used to perform intensive care monitoring.

Concerning integer-valued time series, we may refer the reader to [9, 16, 17, 19]. For a review of various models and their statistical properties, we do recommend [9] where some extensions of integer autoregressive and moving average models are also presented. Many models encountered in the literature are based on thinning operators as defined in [22]. In this paper, we use a more general definition.

Definition 1.1. (*Signed thinning operator.*) Let $\boldsymbol{Y} = \{Y_j\}_{j\in\mathbb{Z}}$ be a sequence of independent and identically distributed (i.i.d.) non-negative integer-valued random variables with mean $\alpha \geqslant 0$ independent of an integer-valued variable X. The thinning operator,

---

* Postal address: Université de Grenoble et CNRS, Laboratoire Jean-Kuntzmann : Équipe Méthodologie Statistique et Sciences Sociales, BP 53, 38041 Grenoble Cédex 9, France
** Postal address: Centre de Recherche en Économie et Statistiques, 15 Boulevard Gabriel Peri, 92245 Malakoff Cédex, France

$\alpha\circ$ is defined by:

$$\alpha \circ X = \begin{cases} \text{sign}(X)\displaystyle\sum_{i=1}^{|X|} Y_i, & \text{if } X \neq 0; \\ 0, & \text{otherwise.} \end{cases}$$

The sequence $\{Y_i\}_{i\in\mathbb{Z}}$ is referred to as a counting sequence. This definition is more general than the usual one where Y is a sequence of Bernoulli random variables with expected value α. (See, for example, [6].) Here, it is a sequence of i.i.d. non-negative integer-valued variables, for example, a sequence of Poisson distributed variables Y_i with parameter α. In fact, any non-negative integer-valued random sequence can be used as a counting series. More, X can take negative values.

To avoid any confusion, if necessary, we can denote the operator by $\alpha(\mathbf{Y})\circ$ or $\alpha(\boldsymbol{\theta})\circ$ instead of $\alpha\circ$ to clearly indicate that it is based on the sequence $\mathbf{Y}$ or that it depends on the parametric vector $\boldsymbol{\theta}$ of the distribution of the variables involved in the operator. Nevertheless, we prefer the simplest notation.

The reader should bear in mind that in Definition 1.1, the mean of the summands Y_i associated with the operator $\alpha\circ$ is α. Suppose $\tilde{\alpha}\circ$ is another thinning operator based on a counting sequence $\{\widetilde{Y}_i\}_{i\in\mathbb{Z}}$. The operators $\alpha\circ$ and $\tilde{\alpha}\circ$ are said to be independent if, and only if, the counting sequences $\{Y_i\}_{i\in\mathbb{Z}}$ and $\{\widetilde{Y}_i\}_{i\in\mathbb{Z}}$ are mutually independent.

Example 1.1. (*Branching process with immigration.*) The Bienaymé-Galton-Watson (BGW) process with immigration can be written using a thinning operator. With this notation, if the offspring of an individual is distributed as Y, and if ζ_t is the immigration contribution to the population at the tth generation, then the classical BGW process satisfies

$$X_t = \alpha \circ X_{t-1} + \zeta_t. \tag{1.1}$$

For each generation t, we need a counting sequence $\mathbf{Y}_t$, so $\{\mathbf{Y}_t\}_{t\in\mathbb{Z}}$ is an i.i.d. process of i.i.d. sequences $\{Y_{t,j}\}_{j\in\mathbb{N}}$. In the case of a BGW process, X_t is never negative. The links between branching processes with immigration and INARMA(p,q) is clearly identified and explained in [2]. □

Example 1.2. (*Inventory monitoring.*) Suppose X_t represents the number of widgets remaining in a distributor inventory at the end of a month. Also suppose if the distributor runs out of stock, he registers the customer order to send it as soon as the widget becomes available. In that case the number of items left at the end of the month could be negative. □

Example 1.3. Given two counting processes, $\{X_t\}$ and $\{Y_t\}$, in some situations we may be interested in the difference between the two processes: $Z_t = X_t - Y_t$, $t \in \mathbb{Z}$, is the excess of X_t over Y_t. Clearly, Z_t can be negative. □

It is clear that in many situations, standard univariate models are not appropriate in the context of integer-valued time series analysis. Using classical real-valued models is even more critical when we cope with a low frequency count data. This has been pinpointed by [21] and many more authors (see [11, 20]). It could explain why integer-valued processes are an important topic and why there have been so many papers on the subject for more than twenty-five years. Many authors use thinning

operators to define integer-valued process similar to classical econometric models. See, for examples, [2, 3, 6, 8, 13].

In [21], a worthy discussion is made on integer-valued ARMA(p, q) processes. In the latter paper, an efficient MCMC algorithm is presented for a wide class of integer-valued autoregressive moving-average processes. In many papers, p and q are assumed to be known. In [7], efficient order selection algorithms are studied for these integer-valued ARMA processes.

It is clear that integer-valued ARMA processes cannot satisfy all practitioner expectations. A common working hypothesis is that the observed time series comes from a stationary process. In some situations, there are good reasons to doubt about this hypothesis.

For example, in Figure 1, we give X_t, the number of campylobacteriosis cases in the Northern Québec, starting in January 1990, with an observation every 28 days.
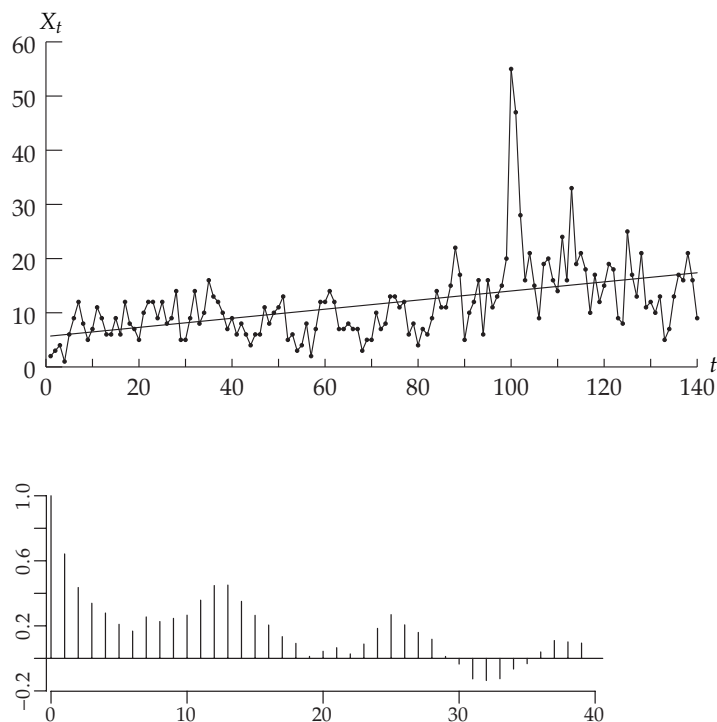
FIGURE 1: Number of campylobacteriosis cases in the Northern Québec, starting in January 1990, 13 regular observations per year. To the top, is the graphic of the original series, to the bottom, is the sample simple correlogram.

One may believe that $E[X_t]$ increases with t. Also, perhaps that there is a structural change happening in the neighborhood of the 100th observation. In [10], problems

---

Infection with a Campylobacter species is one of the most common causes of human bacterial gastroenteritis.

with this series are clearly identified.

For reasons that are similar to the ones we met when we tackle the problem of modeling real valued time series, we have to develop well-adapted tools for practitioner needs. A Dickey-Fuller unit-root type test has been studied by [14]. For a GARCH type model, [8] suggested a process with Poisson conditional distribution with mean and variance λ_t. In [3], the authors tackled the problem of an integer-valued bilinear process. They restricted their works to the following model:

$$X_t = a \circ X_{t-1} + b \circ (\varepsilon_{t-1} X_{t-1}) + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is an i.i.d. sequence of non-negative integer-valued random variables. They proved the existence of this stochastic process, suggested appropriate estimators under a Poissonian hypothesis and applied it to a social medicine series. Recently, [5] cleverly proved the existence of a more general version of this process and inspired in this paper our existence proof of another process (see (3.2)).

The paper has the following structure. In Section 2, we recall a result from [4] giving conditions for the existence of a solution to a quite general model equation in which X_t is expressed in terms of its own past values and the present and past values of a sequence of i.i.d. random variables (*cf.* (2.3)). A quite simple approximation $\{X_t^{(n)}\}$ to $\{X_t\}$ is also given. For this approximation we have:

$$X_t^{(n)} \xrightarrow[n \to \infty]{\mathbb{L}^m} X_t \quad \text{and} \quad X_t^{(n)} \xrightarrow[n \to \infty]{\text{a.s.}} X_t.$$

In Section 3, we give some basic properties of Definition 1.1 thinning operators. Then, two models are presented: the INLARCH model and an integer-valued bilinear type model. Simple conditions for the existence of these processes are given.

Section 4 is devoted to estimation of the parameters. The problem is tackled using a quasi-maximum likelihood estimator for the bilinear model parameters. Before announcing the properties of the estimator, working assumptions and hypotheses are enunciated. Theorem 4.1 claims the strong consistency of the estimators and Theorem 4.2 gives its asymptotic distribution.

In Section 5, we comment consequences of the results when we consider the *almost* classical GINAR(p) process. Proofs are postponed to Section 6.

2. The model

From now on, the sequence $\{\xi_t\}_{t \in \mathbb{Z}}$ is i.i.d. and takes values in a space E' (in many cases E' is just $\mathbb{R}^\infty$). Let $(E, \|\cdot\|)$ be a Banach space. For a random variable $Z \in E$ and a real number $m \geqslant 1$, the expression $\|Z\|_m$ stands for $\left(\mathrm{E}\left[\|Z\|^m\right]\right)^{1/m}$ and $E^{(\mathbb{N})}$, a subset of $E^{\mathbb{N}}$, denotes the set of sequences in $E^{\mathbb{N}}$ with a finite number of non-null terms. Let $F : E^{(\mathbb{N})} \times E' \to E$ be a measurable function and assume there exists a sequence of functions $\{a_j\}_{j \in \mathbb{N}}$ such that for all $\{x_j\}_{j \in \mathbb{N}}$ and $\{y_j\}_{j \in \mathbb{N}}$ in $E^{(\mathbb{N})}$,

$$\|F(0, 0, \ldots; \xi_0)\|_m < +\infty, \tag{2.1a}$$

$$\|F(x_1, x_2, \ldots; \xi_0) - F(y_1, y_2, \ldots; \xi_0)\|_m \leqslant \sum_{j=1}^{\infty} a_j \|x_j - y_j\|, \tag{2.1b}$$

with

$$\sum_{j=1}^{\infty} a_j := a < 1. \tag{2.2}$$

Let us recall a general result of [4] about existence and approximation in $\mathbb{L}^m$ of a stationary process $\{X_t\}_{t\in\mathbb{Z}}$, solution of (2.3):

$$X_t = F\left(X_{t-1}, X_{t-2}, \dots; \xi_t\right). \tag{2.3}$$

The following theorem is a consequence of Theorem 3 and Lemma 6 of [4].

Theorem 2.1. *Assume properties (2.1a) and (2.1b) hold for some $m \geqslant 1$, then there exists a unique stationary solution of (2.3) such that*

$$X_t \in \sigma\left(\xi_t, \xi_{t-1}, \dots\right), \qquad t \in \mathbb{Z}. \tag{2.4}$$

Moreover, the sequence of stationary processes defined $\forall t \in \mathbb{Z}$ as

$$X_t^{(n)} = \begin{cases} F(0; \xi_0), & n = 0; \\ F(X_{t-1}^{(n-1)}, X_{t-2}^{(n-1)}, \dots; \xi_t), & n \geqslant 1; \end{cases}$$

satisfies

$$X_t^{(n)} \xrightarrow[n\to\infty]{\mathbb{L}^m} X_t \quad and \quad X_t^{(n)} \xrightarrow[n\to\infty]{\text{a.s.}} X_t.$$

Remark 2.1. A solution of (2.3) which satisfies (2.4) is always ergodic. Indeed from (2.4), we have:

$$\bigcap_{t\in\mathbb{Z}} \sigma\left(X_{t-1}, X_{t-2}\dots\right) \subset \bigcap_{t\in\mathbb{Z}} \sigma\left(\xi_{t-1}, \xi_{t-2}\dots\right) \tag{2.5}$$

As ξ is i.i.d, any event in the σ-field of the right-hand side of (2.5) has probability 0 or 1 from which we conclude that any event in the σ-field of the left-hand side is also of probability 0 or 1. This shows that the process $\{X_t\}_{t\in\mathbb{Z}}$ is ergodic. The argument comes from [6].

In the sequel, a solution (2.3) satisfying (2.4) will be called a causal solution in $\mathbb{L}^m$. Note that a such solution implies the independence of the σ-algebras $\sigma(X_u : u \leq s)$ and $\sigma(\xi_v : v \geq t)$ when $t > s$.

3. Construction of integer-valued models

3.1. Basic properties of signed thinning operators

Lemma 3.1. *Lets X, Z two random variables and Y, $\widetilde{Y}$ two counting sequences associated with the operators $\alpha\circ$ and $\tilde\alpha\circ$, respectively. Suppose the variance of the counting sequence variables are β and $\tilde\beta$, respectively. Assume that (X, Z), Y, $\widetilde{Y}$ are independent. Let $m \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$. Then:*

1. $\mathrm{E}[\alpha \circ X] = \alpha\mathrm{E}[X]$ *and* $\mathrm{E}\left[(\alpha \circ X)^2\right] = \beta\mathrm{E}[|X|] + \alpha^2\mathrm{E}\left[X^2\right]$.

2. $\mathrm{E}[(\alpha \circ X)(\tilde\alpha \circ Z)] = \alpha\tilde\alpha\mathrm{E}[XZ]$ *and* $\mathrm{cov}[\alpha \circ X, \tilde\alpha \circ Z] = \alpha\tilde\alpha\,\mathrm{cov}[X, Z]$.

3. $\|\alpha \circ X - \alpha \circ Z\|_m \leq \|Y\|_m \|X - Z\|_m$.

4. *For $\ell \geq 2$, we have*

$$\|\alpha \circ X\|_\ell \leq |\alpha| \, \|X\|_\ell + c_\ell \, \|Y - \alpha\|_\ell \, \|X\|_{\ell-1}^{1/2}$$

where the constant $c_\ell > 0$ only depends on ℓ.

Remark 1. Consider the simple model:

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t. \tag{3.1}$$

For $t \in \mathbb{Z}$, let $\boldsymbol{\xi}_t = (\{Y_{t,i}\}_{i \in \mathbb{N}^*}, \varepsilon_t)$. For $x \in \mathbb{Z}$, we define:

$$F(x, \boldsymbol{\xi}_0) = \alpha \circ x_1 + \varepsilon_0.$$

Suppose $(\boldsymbol{\xi}_t)_t$ is an i.i.d. sequence and let $m = 2$. From the result 3. of Lemma 3.1, one has:

$$\|F(x; \boldsymbol{\xi}_0) - F(y; \boldsymbol{\xi}_0)\|_2 \leq \|Y\|_2 \, |x - y|.$$

Moreover if $F(0; \boldsymbol{\xi}_0) = \varepsilon_0 \in \mathbb{L}^2$, we can apply Theorem 2.1 if $\|Y\|_2 < 1$. But this is not optimal. Indeed, it is well known that the condition $\alpha < 1$ is sufficient for the existence and uniqueness in $\mathbb{L}^m$ of a stationary solution of (3.1) (see [18]).

This is the reason why the construction of the model is in two steps. Firstly, we apply Theorem 2.1 with $m = 1$ and get a solution in $\mathbb{L}^1$. Then we use a contraction condition on the means of the counting sequences. Secondly, we show that this solution is still unique in $\mathbb{L}^m$, m being an integer.

3.2. Bilinear model

Let $\{X_t\}_{t \in \mathbb{Z}}$ be a solution to the equation:

$$X_t = \sum_{j=1}^{\infty} \alpha_j \circ X_{t-j} + \varepsilon_t \left(\sum_{j=1}^{\infty} \beta_j \circ X_{t-j} \right) + \eta_t, \tag{3.2}$$

where η_t and ε_t are integer valued random variables in $E = \mathbb{Z}$, $\alpha_j \circ$ and $\beta_j \circ$ being signed thinning operators associated with counting sequences $Y^{(j)}$ and $\widetilde{Y}^{(j)}$ respectively. Theorem 3.1 gives conditions for the existence of a solution to (3.2).

Suppose $\mathrm{E}[\varepsilon_t] = 0$ and for each $t \in \mathbb{Z}$, let:

$$\boldsymbol{\xi}_t = \left(\{Y_{t,i}^{(j)}\}_{(i,j) \in \mathbb{N}^* \times \mathbb{N}^*}, \{\widetilde{Y}_{t,i}^{(j)}\}_{(i,j) \in \mathbb{N}^* \times \mathbb{N}^*}, \varepsilon_t, \eta_t \right).$$

The random variable $\boldsymbol{\xi}_t$ takes values in $\mathbb{Z}^{\mathbb{N}^* \times \mathbb{N}^*} \times \mathbb{Z}^{\mathbb{N}^* \times \mathbb{N}^*} \times \mathbb{Z} \times \mathbb{Z}$. We suppose the process $\{\boldsymbol{\xi}_t\}_{t \in \mathbb{Z}}$ is i.i.d.

Theorem 3.1. *Suppose for an integer $m \geq 1$,*

$$a = \sum_{j=1}^{\infty} \|Y^{(j)}\|_1 + \|\varepsilon\|_m \|\widetilde{Y}^{(j)}\|_1 < 1, \qquad \sum_{j=1}^{\infty} \|Y^{(j)}\|_m + \|\widetilde{Y}^{(j)}\|_m + \|\eta_0\|_m < \infty, \tag{3.3}$$

then there exists a unique causal solution to (3.2) in $\mathbb{L}^m$.

3.3. INLARCH(∞) time series model

An INLARCH(∞) time series model satisfies

$$X_t = \alpha \circ \varepsilon_t + \sum_{j=1}^{\infty} \alpha_j \circ (\varepsilon_t X_{t-j}), \quad t \in \mathbb{Z}. \tag{3.4}$$

For $j \in \mathbb{N}^*$, we will denote by $Y^{(j)}$ (resp. Y) the counting sequences associated with the operator $\alpha_j \circ$ (resp. $\alpha \circ$).

As for the bilinear model, we suppose $\{\xi_t\}_{t \in \mathbb{Z}}$ is an i.i.d. sequence. Theorem 3.2 states a sufficient condition for the existence of this process.

Theorem 3.2. *Suppose for an integer $m \geq 1$,*

$$a = \|\varepsilon\|_m \sum_{j \in \mathbb{N}^*} \|Y^{(j)}\|_1 < 1, \qquad \sum_{j \in \mathbb{N}^*} \|Y^{(j)}\|_m + \|Y\|_m < \infty,$$

then equation (3.4) admits a unique causal solution in $\mathbb{L}^m$.

4. Quasi-maximum likelihood estimator in bilinear model

This section aims at giving a quasi-maximum likelihood estimators (QMLE) for the parameters of the bilinear model (3.2) with a finite number of terms in the two summations. Without lost of generality, we may assume that there are p terms in each summation; otherwise some $\alpha_j \circ$ or $\beta_j \circ$ are $0 \circ$. The equation satisfied by the process is:

$$X_t = \sum_{j=1}^{p} \alpha_j \circ X_{t-j} + \varepsilon_t \sum_{j=1}^{p} \beta_j \circ X_{t-j} + \eta_t. \tag{4.1}$$

Example 4.1. Let $p = 2$ and consider

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-1} + \varepsilon_t \beta_1 \circ X_{t-1} + \eta_t,$$

where: $\alpha_1 \circ$ is based on a Bernoulli counting series with $p = 1/2$; $\alpha_2 \circ$ and $\beta_1 \circ$, on Poisson counting series with means $1/8$ and $1/2$, respectively; $\{\eta_t\}$ is a sequence of i.i.d. Poisson random variables with parameter $\lambda = 1/2$; $\{\varepsilon_t\}$ is a sequence of differences between two independent Poisson variables with parameter $\lambda = 2/3$. A simulated trajectory is presented in Figure 2. Note that there is "period", just after $t = 80$ with quite high values compared to the other ones. □

For $(t, j) \in \mathbb{Z} \times \{1, \ldots, p\}$, we define the following $\sigma-$algebras:

$$\mathcal{F}_t = \sigma(X_{t-k} : k \in \mathbb{N}^*), \quad \mathcal{G}_{t,j} = \sigma(Y_{t,i}^{(j)} : i \in \mathbb{N}^*), \quad \text{and} \quad \widetilde{\mathcal{G}}_{t,j} = \sigma(\widetilde{Y}_{t,i}^{(j)} : i \in \mathbb{N}^*),$$

From now on, we suppose the following working assumptions are satisfied:

1. $\{\xi_t\}_{t \in \mathbb{Z}}$ is an i.i.d. sequence of random variables.

2. For all $t \in \mathbb{Z}$, the σ-algebras $\mathcal{G}_{t,1}, \ldots, \mathcal{G}_{t,p}$ (resp. $\widetilde{\mathcal{G}}_{t,1}, \ldots, \widetilde{\mathcal{G}}_{t,p}$) are independent.
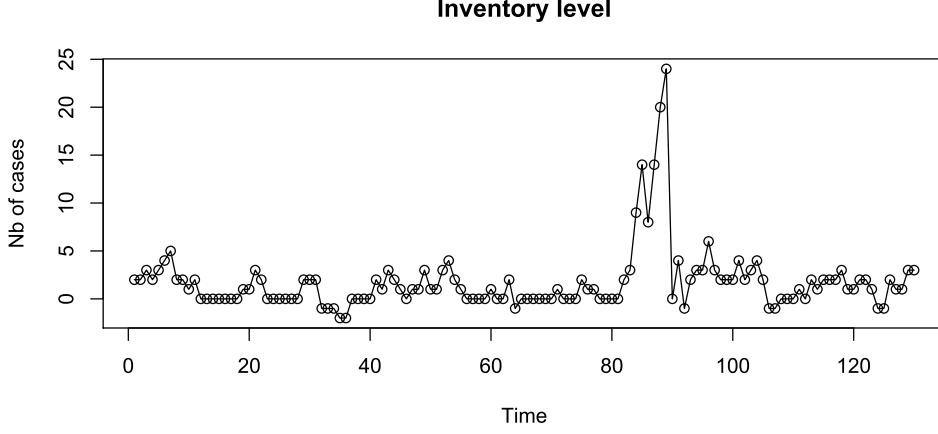
Inventory level

FIGURE 2: Simulated trajectory generated by model $X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-1} + \varepsilon_t \beta_1 \circ X_{t-1} + \eta_t$, where: $\alpha_1 \circ$ is based on a Bernoulli counting series with $p = 1/2$; $\alpha_2 \circ$ and $\beta_1 \circ$, on Poisson counting series with means $1/8$ and $1/2$, respectively; $\{\eta_t\}$ is a sequence of i.i.d. Poisson random variables with parameter $\lambda = 1/2$; $\{\varepsilon_t\}$ is a sequence of differences between two independent Poisson variables with parameter $\lambda = 2/3$.

3. For all $t \in \mathbb{Z}$, the σ-algebras $\sigma(\varepsilon_t)$, $\sigma(\eta_t)$ and $(\vee_{1 \leq j \leq p} \mathcal{G}_{t,j}) \vee (\vee_{1 \leq j \leq p} \widetilde{\mathcal{G}}_{t,j})$ are mutually independent.

Remark 4.1. Assumptions with respect to the σ-algebras allow dependence between the set of operators $\{\alpha_j \circ\}_{1 \leqslant j \leqslant p}$ and the set of operators $\{\beta_j \circ\}_{1 \leqslant j \leqslant p}$.

For an integer $d \geq 1$, let Θ be a subset of $\mathbb{R}^d$ and $\theta_0 \in \Theta$. For $1 \leq j \leq p$, consider functions $b_j, c_j, w_j, \mu, \nu : \Theta \to \mathbb{R}$ such that:

i) $b_j(\theta_0) = \alpha_j$ and $c_j(\theta_0) = \beta_j$.
   To ensure identifiability, we suppose there exists $j_0 \in \{1, \ldots, p\}$ such that $\beta_{j_0} > 0$ and the function c_{j_0} is positive on Θ.

ii) $w_j(\theta_0) = \text{var}\left[Y^{(j)}\right] + \sigma^2 \times \text{var}\left[\widetilde{Y}^{(j)}\right]$, $\sigma^2 = \text{var}[\varepsilon]$.

iii) $\mu(\theta_0) = \text{E}[\eta]$ and $\nu(\theta_0) = \text{var}[\eta]$.

The following hypotheses will also be required.

H1) Θ is a compact subset of $\mathbb{R}^d$.

H2) Condition (3.3) holds with $m = 2$.

H3) The distribution support of η_t contains at least 5 different points if $\text{var}[\varepsilon_t] \neq 0$ and 3, otherwise.

H4) The following condition is satisfied : $h = \inf_{\theta \in \Theta} \nu(\theta) > 0$.

H5) The function $f : \Theta \to \mathbb{R}^{3p+2}$ defined by

$$f(\boldsymbol{\theta}) = \left((b_j \boldsymbol{\theta}), c_j(\boldsymbol{\theta}), w_j(\boldsymbol{\theta}))_{1 \le j \le p}, \mu(\theta), \nu(\theta) \right)$$

is injective and continuous on $\boldsymbol{\Theta}$.

For $(t, \boldsymbol{\theta}) \in \mathbb{Z} \times \boldsymbol{\Theta}$, let

$$m_t(\boldsymbol{\theta}) = \mu(\boldsymbol{\theta}) + \sum_{j=1}^{p} b_j(\boldsymbol{\theta}) X_{t-j}$$

and

$$V_t(\boldsymbol{\theta}) = \sigma^2 \left(\sum_{j=1}^{p} c_j(\boldsymbol{\theta}) X_{t-j} \right)^2 + \sum_{j=1}^{p} w_j(\boldsymbol{\theta}) \left| X_{t-j} \right| + \nu(\boldsymbol{\theta}).$$

Observe that under assumption H4, we have:

$$\inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} V_t(\boldsymbol{\theta}) \ge h, \qquad \text{a.s.} \tag{4.2}$$

Lemma 4.1. *Let $\{X_t\}$ given by (4.1). We have:*

$$\mathrm{E}[X_t \mid \mathcal{F}_{t-1}] = m_t(\boldsymbol{\theta}_0), \quad \mathrm{var}[X_t \mid \mathcal{F}_{t-1}] = V_t(\boldsymbol{\theta}_0).$$

Remark 4.2. On the one hand, the conditional expectation is the same as the one of a GINAR(p) process. On the other hand, a second-degree polynomial appears in the conditional variance.

4.1. Estimators definition

For the estimation of the parameters, no distribution assumptions are made and a quasi-maximum likelihood approach turns out to be well suited to this setup. The maximum is found assuming a conditional Gaussian density for X_t, given the past until time $t-1$. In [23] this method is used in ARCH modeling.

Let us give the details for model (4.1). Suppose we observe $X_{-p+1}, \ldots, X_0$ and let:

$$q_t(\boldsymbol{\theta}) = \frac{(X_t - m_t(\boldsymbol{\theta}))^2}{V_t(\boldsymbol{\theta})} + \ln V_t(\boldsymbol{\theta}), \qquad t \ge 1;$$

$$Q_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} q_t(\boldsymbol{\theta});$$

$$Q(\boldsymbol{\theta}) = \mathrm{E} \left[\left(\frac{(X_0 - m_0(\boldsymbol{\theta}))^2}{V_0(\boldsymbol{\theta})} + \ln V_0(\boldsymbol{\theta}) \right) \right];$$

$$\hat{\boldsymbol{\theta}}_T = \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q_T(\boldsymbol{\theta}).$$

So, $\hat{\boldsymbol{\theta}}_T$ is the QMLE for $\boldsymbol{\theta}$ and the actual value of $\boldsymbol{\theta}$ is $\boldsymbol{\theta}_0$.

4.1.1. Consistency of the estimator.

Theorem 4.1. *Under hypotheses H1 to H5, the estimator $\hat{\boldsymbol{\theta}}_T$ is a strongly consistent estimator of $\boldsymbol{\theta}_0$: $\hat{\boldsymbol{\theta}}_T \xrightarrow[n \to \infty]{a.s.} \boldsymbol{\theta}_0$.*

4.1.2. Asymptotic normality of QMLE. In the following, if g is a function, $g : \Theta \mapsto \mathbb{R}$, ∇g is its gradient and $\nabla^2 g$ is its Hessian matrix.

Other hypotheses are needed.

H7) Condition (3.3) holds with $m = 4$.

H8) The f function is twice differentiable on Θ and rank $\nabla f(\boldsymbol{\theta}_0) = d$. More, $\inf_{\boldsymbol{\theta} \in \Theta} w_j(\boldsymbol{\theta}) > 0, \forall j = 1, \dots, p$.

H9) $\boldsymbol{\theta}_0$, the actual value of $\boldsymbol{\theta}$, is an interior point of Θ, $\boldsymbol{\theta}_0$, that is $\boldsymbol{\theta}_0 \in \Theta^\circ$.

Theorem 4.2. *Under hypotheses H1, $\dots$, H9, the estimator $\hat{\boldsymbol{\theta}}_T$ is asymptotically normal:*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow[T \to \infty]{L} \mathcal{N}(0, F_0^{-1} G_0 F_0^{-1})$$

where

$$
\begin{aligned}
F_0 &= \mathrm{E}\left[\nabla^2 q_0(\boldsymbol{\theta}_0)\right] \\
&= \mathrm{E}\left[V_0(\theta_0)^{-2} \nabla V_0(\theta_0) \nabla V_0(\theta_0)^\top\right] + 2\mathrm{E}\left[V_0(\theta_0)^{-1} \nabla m_0(\theta_0) \nabla m_0(\theta_0)^\top\right]
\end{aligned}
$$

and

$$
\begin{aligned}
G_0 &= \mathrm{var}[\nabla q_0(\boldsymbol{\theta}_0)] \\
&= \mathrm{E}\left[V_0(\theta_0)^{-4}(X_0 - m_0(\theta_0))^4 \nabla V_0(\theta_0) \nabla V_0(\theta_0)^\top\right] \\
&\quad - \mathrm{E}\left[V_0(\theta_0)^{-2} \nabla V_0(\theta_0) \nabla V_0(\theta_0)^\top\right] + 4\mathrm{E}\left[V_0(\theta_0)^{-1} \nabla m_0(\theta_0) \nabla m_0(\theta_0)^\top\right] \\
&\quad + \mathrm{E}\left[V_0(\theta_0)^{-3}(X_0 - m_0(\theta_0))^3 \nabla V_0(\theta_0) \nabla m_0(\theta_0)^\top\right] \\
&\quad + \mathrm{E}\left[V_0(\theta_0)^{-3}(X_0 - m_0(\theta_0))^3 \nabla m_0(\theta_0) \nabla V_0(\theta_0)^\top\right]
\end{aligned}
$$

5. QMLE for GINAR(p) processes

When $\sigma^2 = 0$, (4.1) leads to a GINAR(p) process:

$$X_t = \sum_{j=1}^p \alpha_j \circ X_{t-j} + \eta_t.$$

Estimation for this process has been tackled by least squares (see [6, 18]).

The conditional least squares estimator is given by:

$$\hat{\boldsymbol{\theta}}_T = \text{argmin}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{T} \sum_{t=1}^{T} (X_t - m_t(\boldsymbol{\theta}))^2 .$$

But this approach cannot be applied to obtain estimators for all the parameters if the probability distribution of the counting sequences depends on two parameters or more.

In particular, suppose the operators $\alpha_j \circ$, $1 \leqslant j \leqslant p$, are counting series with variables for which the support is a 3-point set $\{a, b, c\}$ and we want to estimate:

$$(q_{a,j}, q_{b,j}) = \left(\Pr\left(Y_{0,0}^{(j)} = a \right), \Pr\left(Y_{0,0}^{(j)} = b \right) \right)$$

$1 \leq j \leq p$, as well as $(\text{E}[\eta], \text{var}[\eta])$. Let

$$\boldsymbol{\theta}_0 = \left(q_{a,1}, q_{b,1}, \ldots, q_{a,p}, q_{b,p}, \text{E}[\eta_0], \text{var}[\eta_0] \right) \in \boldsymbol{\Theta} \subset \mathbb{R}^{2p+2}.$$

For $1 \leq j \leq p$, let

$$b_j(\boldsymbol{\theta}) = (a - c)\boldsymbol{\theta}_{2j-1} + (b - c)\boldsymbol{\theta}_{2j} + c$$

$$w_j(\boldsymbol{\theta}) = (a^2 - c^2)\boldsymbol{\theta}_{2j-1} + (b^2 - c^2)\boldsymbol{\theta}_{2j} + c^2 - b_j(\boldsymbol{\theta})^2$$

$$\mu(\boldsymbol{\theta}) = \boldsymbol{\theta}_{2p+1}, \qquad \nu(\boldsymbol{\theta}) = \boldsymbol{\theta}_{2p+2}$$

A least squares approach is not tractable because $\boldsymbol{\theta}_0$ is not identifiable by just considering $m_t(\boldsymbol{\theta}) = \sum_{j=1}^{p} b_j(\boldsymbol{\theta}) X_{t-j}$, $t \in \mathbb{Z}$, the conditional means of the process $\{X_t\}$. In fact, the function $\boldsymbol{\theta} \mapsto \left(b_1(\boldsymbol{\theta}), \ldots, b_p(\boldsymbol{\theta}), \mu(\boldsymbol{\theta}), \nu(\boldsymbol{\theta}) \right)$ is not injective. However, it is clear that the function:

$$\boldsymbol{\theta} \mapsto \left(b_1(\boldsymbol{\theta}), w_1(\boldsymbol{\theta}), \ldots, b_p(\boldsymbol{\theta}), w_p(\boldsymbol{\theta}), \mu(\boldsymbol{\theta}), \nu(\boldsymbol{\theta}) \right)$$

is injective and we can use Section 4.1 results to estimate parameter $\boldsymbol{\theta}_0$.

Example 5.1. Let us return to Example 4.1. It is quite easy to proceed to the estimation of the parameters using a widespread and simple tool like Microsoft Excel. We use the Excel's Solver macro to find the optimum. The estimated model is:

$$X_t = \hat{\alpha}_1 \circ X_{t-1} + \hat{\alpha}_2 \circ X_{t-1} + \varepsilon_t \hat{\beta}_1 \circ X_{t-1} + \eta_t,$$

where: $\hat{\alpha}_1 \circ$ is based on a Bernoulli counting series with $\hat{p} = 0.65$; $\hat{\alpha}_2 \circ$ and $\hat{\beta}_1 \circ$, on Poisson counting series with means 0.12 and 0.58 respectively; $\{\eta_t\}$ a sequence of i.i.d. Poisson random variables with parameter $\lambda = 0.47$; $\{\varepsilon_t\}$ is a sequence of differences between two independent Poisson variables with parameter $\lambda = 0.49$. So, $\hat{\boldsymbol{\theta}} = (0.645; 0.120, 0.503; 0.669; 0.469)^\top$. Recall that the actual value of the parameter is : $\boldsymbol{\theta} = (0.5; 0.125; 0.5; 0.66\overline{7}; 0.500)^\top$.

6. Extended proofs of the results

6.1. Proof of Lemma 3.1

Proof.   1. $\mathrm{E}_{X=x}[\alpha \circ X] = \mathrm{E}\left[\mathrm{sign}(x) \sum_{i=1}^{|x|} Y_i\right] = x\alpha$ and the first result follows from expectation with respect to X.

For the second point, note that:

$$\mathrm{E}_{X=x}\left[(\alpha \circ X)^2\right] = \mathrm{E}\left[\sum_{i=1}^{|x|} Y_i\right]^2 = |x|\,\mathrm{E}[Y^2] + |x|\,(|x|-1)\alpha^2,$$

and again the result follows from expectation with respect to X.

2. Since the variables $(\alpha \circ X, Z)$ and $\widetilde{Y}$ are independent, from result 1., we get the following equality:

$$\mathrm{E}[(\alpha \circ X) \times (\widetilde{\alpha} \circ Z)] = \widetilde{\alpha}\mathrm{E}[(\alpha \circ X) \cdot Z)].$$

As Y is independent of (X, Z), we obtain $\widetilde{\alpha}\mathrm{E}[(\alpha \circ X) \cdot Z)] = \alpha\widetilde{\alpha}\mathrm{E}[XZ]$. The second assertion is obvious.

3. We use the first point of item 1 and if $x, z \in \mathbb{Z}$:

$$\|\alpha \circ x - \alpha \circ z\|_m \le \|Y\|_m\,|x - z|.$$

Independence between Y and (X, Z) yields the result after expectation with repesct to X and Z.

4. See [5, Theorem 2.2], for a proof of this inequality.

$\square$

6.2. Proof of Theorem 3.1

The demonstration proceeds in two steps. Firstly, we show that under Theorem 3.1 hypotheses, equation (3.2) has a unique causal solution in $\mathbb{L}^1$. Then, we show that this solution has moments of order m.

To show the existence in $\mathbb{L}^1$, we use Theorem 2.1. Let $F : \mathbb{Z}^{(\mathbb{N}^*)} \times \mathbb{Z} \to \mathbb{Z}$ be:

$$F\left(\{x_j\}_{j \in \mathbb{N}^*}; \xi_0\right) = \sum_{j=1}^{\infty} \alpha_j \circ x_i + \varepsilon_0 \left(\sum_{j=1}^{\infty} \beta_j \circ x_i\right) + \eta_0.$$

We have: $\|F(0; \xi_0)\|_1 = \|\eta_0\|_1 < \infty$. More, by the result 3. of Lemma 3.1, we get:

$$\left\|F\left(\{x_j\}_{j \in \mathbb{N}^*}; \xi_0\right) - F\left(\{y_j\}_{j \in \mathbb{N}^*}; \xi_0\right)\right\|_1 \le \sum_{j=1}^{\infty} \left(\|Y^{(j)}\|_1 + \|\varepsilon_0\|_1 \|\widetilde{Y}^{(j)}\|_1\right)|x_j - y_j|$$

$$\le \sum_{j=1}^{\infty} \left(\|Y^{(j)}\|_1 + \|\varepsilon_0\|_m \|\widetilde{Y}^{(j)}\|_1\right)|x_j - y_j|.$$

Because $a = \sum_{j=1}^{\infty} \left(\|Y^{(j)}\|_1 + \|\varepsilon_0\|_m \|\widetilde{Y}^{(j)}\|_1 \right) < 1$, we can apply Theorem 2.1 and conclude that there exists in $\mathbb{L}^1$ a unique causal stationary process $\{X_t\}$, solution to (3.2), such that $\|X_t\|_1 < \infty$.

Let us show that $X_t \in \mathbb{L}^m$. To this end, let us introduce the stationary process defined by:

$$X_{n,t} = \begin{cases} F(0; \boldsymbol{\xi}_t), & n = 0; \\ F\left(\{X_{n-1,t-j}\}_{j\geq 1}; \boldsymbol{\xi}_t\right), & n \geq 1; \end{cases} \quad t \in \mathbb{Z}.$$

By Theorem 3.1, we have:

$$X_{n,t} \xrightarrow[n\to\infty]{\text{a.s.}} X_t \quad \text{and} \quad X_{n,t} \xrightarrow[n\to\infty]{\mathbb{L}^1} X_t.$$

Next, we show that $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_m < \infty$. From this last inequality, using Fatou's Lemma, we will conclude that

$$\|X_0\|_m \leq \liminf_{n\to\infty} \|X_{n,0}\|_m < \infty.$$

We use induction to show that for each $\ell \in \{1, \dots, m\}$, we have $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_\ell < \infty$. Since $\lim_{n\to\infty} X_{n,0} = X_n$ in $\mathbb{L}^1$, we have $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_1 < \infty$ and the result follows for $\ell = 1$.

Suppose for $\ell \in \{1, \dots, m-1\}$ we have $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_\ell < \infty$. We want to show that $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_{\ell+1} < \infty$. Let $n \in \mathbb{N}$. We have:

$$\|X_{n+1,0}\|_{\ell+1} \leqslant \sum_{j=1}^{\infty} \|\alpha_j \circ X_{n,-j}\|_{\ell+1} + \|\varepsilon_0\|_{\ell+1} \sum_{j=1}^{\infty} \|\beta_j \circ X_{n,-j}\|_{\ell+1} + \|\eta_0\|_{\ell+1}.$$

To simplify the equations writing, let :

$$d_{j,h} = \|Y^{(j)} - \alpha_j\|_h + \|\varepsilon_0\|_h \|\widetilde{Y}^{(j)} - \beta_j\|_h, \quad \text{for } j \geq 1 \text{ and } h \in \{1, \dots, m\}.$$

Using result 4. of Lemma 3.1, we get:

$$\begin{aligned} \|X_{n+1,0}\|_{\ell+1} &\leq c_{\ell+1} \sum_{j=1}^{\infty} d_{j,\ell+1} \|X_{n,-j}\|_\ell^{1/2} + \|\eta_0\|_{\ell+1} \\ &\quad + \sum_{j=1}^{\infty} \left(|\alpha_j| + |\beta_j| \|\varepsilon_0\|_{\ell+1} \right) \|X_{n,-j}\|_{\ell+1} \\ &\leq a \|X_{n,0}\|_{\ell+1} + B. \end{aligned}$$

where $B = c_{\ell+1} \sup_k \|X_{k,0}\|_\ell^{1/2} \sum_{j=1}^{\infty} d_{j,l+1} + \|\eta_0\|_{\ell+1}$.
As $X_{0,0} = \eta_0$, this leads to $\|X_{n+1,0}\|_{\ell+1} \leq a^{n+1} \|\eta_0\|_{\ell+1} + B \sum_{i=1}^{n} a^i$.
Observe that

$$B \leq c_{\ell+1} \sup_k \|X_{k,0}\|_\ell^{1/2} \sum_{j=1}^{\infty} d_{j,m} + \|\eta_0\|_m.$$

Then by condition (3.3) and the induction hypothesis, B is finite and we get:

$$\|X_{n+1,0}\|_{\ell+1} \leq \|\eta_0\|_{\ell+1} + \frac{B}{1-a}$$

and $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_{\ell+1} < \infty$. Hence, by finite induction on the subset $\{1,\dots,m\}$, we have $\sup_{n\in\mathbb{N}} \|X_{n,0}\|_m < \infty$. Finally, by the remark made previously, $\|X_0\|_m < \infty$. Uniqueness in $\mathbb{L}^m$ follows from uniqueness in $\mathbb{L}^1$.   □

6.3. Proof of Lemma 4.1

The conditional expectation of X_t given the past until time $t-1$ is :

$$\mathrm{E}[X_t \mid \mathcal{F}_{t-1}] = \mathrm{E}[\eta] + \sum_{j=1}^{p} \mathrm{E}_{\mathcal{F}_{t-1}}[\alpha_j \circ X_{t-j}] = \mathrm{E}[\eta] + \sum_{j=1}^{p} \alpha_j X_{t-j} = m_t(\boldsymbol{\theta}_0).$$

For the conditional variance, we get:

$$\mathrm{var}[X_t \mid \mathcal{F}_{t-1}] = \mathrm{var}_{\mathcal{F}_{t-1}}\left[\sum_{j=1}^{p} \alpha_j \circ X_{t-j}\right] + \sigma^2 \mathrm{E}_{\mathcal{F}_{t-1}}\left[\left(\sum_{j=1}^{p} \beta_j \circ X_{t-j}\right)\right] + \mathrm{var}[\eta]$$

Simple computations lead to

$$\mathrm{E}_{\mathcal{F}_{t-1}}\left[(\alpha_j \circ X_{t-j})(\alpha_k \circ X_{t-k})\right] = \begin{cases} \alpha_j \alpha_k X_{t-j} X_{t-k}, & j \neq k; \\ \alpha_j^2 X_{t-j}^2 + \mathrm{var}\left[Y^{(j)}\right]|X_{t-j}|, & j = k. \end{cases}$$

Similar formulas can be found if $\beta_j\circ$ is substituted for $\alpha_j\circ$. Using these expressions in $\mathrm{var}[X_t \mid \mathcal{F}_{t-1}]$ expansion leads to the final expression:

$$\mathrm{var}[X_t \mid \mathcal{F}_{t-1}] = \sigma^2 \left(\sum_{j=1}^{p} \beta_j X_{t-j}\right)^2 + \sum_{j=1}^{p} w_j(\boldsymbol{\theta}_0)|X_{t-j}| + \mathrm{var}[\eta],$$

where $w_j(\boldsymbol{\theta}_0) = \mathrm{var}[Y^{(j)}] + \sigma^2 \mathrm{var}[\widetilde{Y}^{(j)}]$. This is exactly $V_t(\boldsymbol{\theta}_0)$.   □

6.4. Proof of Theorem 3.2

The proof is very similar to proof given to Theorem 3.1 and is omitted.   □

6.5. Proof of Theorem 4.1

Before giving the demonstration, some intermediate results are required. Let us recall Theorem 6.1 from [23].

Theorem 6.1. *Let Θ a compact set of $\mathbb{R}^d$ and $\{v_t\}_{t\in\mathbb{Z}}$ a stationary ergodic sequence of random elements with values in $\mathbb{C}(\Theta, \mathbb{R})$. Then the uniform strong law of large numbers is implied by*

$$\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\Theta} |v_0(\boldsymbol{\theta})|\right] < \infty.$$

Lemma 6.1 follows from Theorem 6.1.

Lemma 6.1.
$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |Q_T(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})| \xrightarrow[T \to \infty]{} 0.$$

Proof. Let us verify that Theorem 6.1 hypotheses are satisfied. Firstly, we prove that $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\{q_t(\boldsymbol{\theta})\}_{t \geqslant 1}$ is an ergodic stationary sequence. From Remark 2.1, $\{X_t\}$ is a stationary ergodic process. More, for $(t, \boldsymbol{\theta}) \in \mathbb{Z} \times \boldsymbol{\Theta}$, by definition of $q_t(\boldsymbol{\theta})$, there exists a measurable function $f_{\boldsymbol{\theta}}$ defined on $\mathbb{R}^{p+1}$ such that $q_t(\boldsymbol{\theta}) = f_{\boldsymbol{\theta}}(X_t, \dots, X_{t-p})$. This implies that the sequence $\{q_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ is also stationary and ergodic.

Secondly, we have:

$$|q_0(\boldsymbol{\theta})| \leqslant \frac{[X_0 - m_0(\boldsymbol{\theta})]^2}{h} + |\ln(V_0(\boldsymbol{\theta}))| .$$

Then, from H1, H2, H4 and H5, we get:

$$|X_0 - m_0(\boldsymbol{\theta})| \leqslant |X_0| + \|\mu\|_\infty + \sum_{j=1}^{p} \|b_j\|_\infty |X_{-j}| \in \mathbb{L}^2,$$

and

$$h \leqslant V_0(\boldsymbol{\theta}) \leqslant \sigma^2 \left(\sum_{j=1}^{p} \|c_j\|_\infty |X_{-j}| \right)^2 + \sum_{j=1}^{p} \|w_j\|_\infty |X_{-j}| + \|\nu\|_\infty \in \mathbb{L}^1.$$

This shows that $\mathrm{E}\left[\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |q_0(\boldsymbol{\theta})|\right] < \infty$. Moreover, from assumption H5, the function $\boldsymbol{\theta} \mapsto q_0(\boldsymbol{\theta})$ is continuous and Theorem 6.1 leads to the result.

In the sequel, we use $\boldsymbol{Z}_t$ to denote $(X_{t-1}, \dots, X_{t-p})$, $t \in \mathbb{Z}$.

Lemma 6.2. *Let $t \in \mathbb{Z}$. Then for any realization $\{z_t\}$ of $\{\boldsymbol{Z}_t\}$, the distribution support of the random variable $X_{t|\boldsymbol{Z}_t = z_t}$ has at least five points if $\sigma \neq 0$ and at least three, if $\sigma = 0$.*

Proof. The distribution of $X_{t|\boldsymbol{Z}_t = z_t}$ is the same as the distribution of $C_{z_t} + \eta_t$ with

$$C_{z_t} = \sum_{j=1}^{p} \alpha_j \circ x_{t-j} + \varepsilon_t \sum_{j=1}^{p} \beta_j \circ x_{t-j}.$$

By H3 and using the fact that η_t are C_{z_t} independent, the result follows.

Lemma 6.3 will also be required for Theorem 4.1 demonstration.

Lemma 6.3. *Let $t \in \mathbb{Z}$. We have:*

1. *If $\sum_{j=1}^{p} \gamma_j X_{t-j} = \gamma$ then $\gamma = \gamma_j = 0, \forall j \in \{1, \dots, p\}$.*

2. *If we suppose $\sigma \neq 0$ and $\left(\sum_{j=1}^{p} s_j X_{t-j} \right) \left(\sum_{j=1}^{p} u_j X_{t-j} \right) + \sum_{j=1}^{p} \gamma_j |X_{t-j}| = \gamma$, then either $s_j = \gamma_j = \gamma = 0, \forall j \in \{1, \dots, p\}$ or $u_j = \gamma_j = \gamma = 0, \forall j \in \{1, \dots, p\}$.*

3. *If we suppose $\sigma = 0$ and $\sum_{j=1}^{p} \gamma_j |X_{t-j}| = \gamma$, then $\gamma_j = \gamma = 0$, $j = 1, \ldots, p$.*

Proof.   1. suppose $m = \min\{j \in \{1, \ldots, p\} : \gamma_j \neq 0\}$ exists. Then, X_{t-m} is measurable with respect to $\mathcal{F}_{t-m-1}$. This is in contradiction of Lemma 6.2. Hence, we deduce that $\gamma_j = 0$, $\forall j \in \{1, \ldots, p\}$ from which it follows that $\gamma = 0$.

2. Suppose that $m = \min\{j \in \{1, \ldots, p\} : |s_j| + |u_j| + |\gamma_j| \neq 0\}$ exists. Note that if m does not exist, the result is obviously true.

    Suppose first that $m \leq p - 1$. Let

$$F(\mathbf{Z}_{t-m}) = \sum_{j=m+1}^{p} (u_m s_j + s_m u_j) X_{t-j}$$

    and

$$G(\mathbf{Z}_{t-m}) = \gamma - \left(\sum_{j=m+1}^{p} s_j X_{t-j} \right) \left(\sum_{j=m+1}^{p} u_j X_{t-j} \right) - \sum_{j=m+1}^{p} \gamma_j |X_{t-j}|$$

    We have

$$s_m u_m X_{t-m}^2 + F(\mathbf{Z}_{t-m}) X_{t-m} + \gamma_m |X_{t-m}| = G(\mathbf{Z}_{t-m}).$$

    Using Lemma 6.2, we see that for any realization z_{t-m} of $\mathbf{Z}_{t-m}$, there exist five solutions to the equation with unknown x:

$$s_m u_m x^2 + F(z_{t-m}) x + \gamma_m |x| = G(z_{t-m}).$$

    Consequently, $s_m u_m = 0$, $G(z_{t-m}) = 0$ and $|F(z_{t-m})| = |\gamma_m|$.

    Without loss of generality, suppose $s_m = 0$. Then, the random variable $F(\mathbf{Z}_{t-m}) = \sum_{j=m+1}^{p} u_m s_j X_{t-j}$ can take only two values almost surely: $\pm \gamma_m$. If $u_m = 0$, then $\gamma_m = 0$ and this is in contradiction of the assumption that m exists. Hence, $u_m \neq 0$. suppose $r = \min\{j : m + 1 \leq j \leq p, s_j \neq 0\}$ exists. As

$$\sum_{j=r}^{p} s_j X_{t-j} \in \{\gamma_m / u_m, -\gamma_m / u_m\}$$

    we conclude that for any realization z_{t-r} of the random vector $\mathbf{Z}_{t-r}$, the distribution support of the conditional law $X_{t-r|\mathbf{Z}_{t-r}=z_{t-r}}$ has two points. This is in contradiction of Lemma 6.2. Hence, $s_j = 0$, $\forall j \in \{1, \ldots, p\}$. Hence equality $G(\mathbf{Z}_{t-m}) \overset{a.s.}{=} 0$ leads to

$$\sum_{j=m+1}^{p} \gamma_j |X_{t-j}| \overset{a.s.}{=} \gamma.$$

    If $q = \inf\{j \in \{m + 1, \ldots, p\} : \gamma_j \neq 0\}$ exists, the distribution support of the conditional law $X_{t-q|\mathbf{Z}_{t-q}=z_{t-q}}$ contains only two values. This is in contradiction of Lemma 6.2. So, $\gamma_j = 0$, $\forall j \geq 1$. Finally $\gamma = 0$ and the result follows.

In the case where $m = p$, we have $s_p u_p X_{t-p}^2 + \gamma_p |X_{t-p}| = \gamma$. By Lemma 6.2, we conclude that necessarily $s_p u_p = \gamma_p = \gamma = 0$.

3. Suppose $m = \min\{j \in \{1, \ldots, p\} : \gamma_j \neq 0\}$ exists. Then, $|X_{t-m}|$ is measurable with respect to $\mathcal{F}_{t-m-1}$. Hence for each z_{t-m} the distribution support of the conditional law of $X_{t-m \mid Z_{t-m} = z_{t-m}}$ contains at most two points. This is in contradiction of Lemma (6.2) and the result follows.

Lemma 6.4. *If, $m_0(\boldsymbol{\theta}) = m_0(\boldsymbol{\theta}_0)$ and $V_0(\boldsymbol{\theta}) = V_0(\boldsymbol{\theta}_0)$ are satisfied, then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.*

Proof. Let us suppose $m_0(\boldsymbol{\theta}) = m_0(\boldsymbol{\theta}_0)$. Applying the first point of Lemma 6.3 with $\gamma_j = b_j(\boldsymbol{\theta}) - b_j(\boldsymbol{\theta}_0)$, $1 \leq j \leq p$ and $\gamma = \mu(\boldsymbol{\theta}_0) - \mu(\boldsymbol{\theta})$, we obtain $b_j(\boldsymbol{\theta}) = b_j(\boldsymbol{\theta}_0)$, $j = 1, \ldots, p$ and $\mu(\boldsymbol{\theta}) = \mu(\boldsymbol{\theta}_0)$.
More, suppose $V_0(\boldsymbol{\theta}) = V_0(\boldsymbol{\theta}_0)$. Two cases need to be considered.

- Firstly, assume that $\sigma^2 = \text{var}[\varepsilon_t] \neq 0$. We apply result 2. of Lemma 6.3 setting for $j \in \{1, \ldots, p\}$:

$$s_j = \sigma(c_j(\boldsymbol{\theta}) - c_j(\boldsymbol{\theta}_0)), \quad u_j = \sigma(c_j(\boldsymbol{\theta}) + c_j(\boldsymbol{\theta}_0)), \quad \gamma_j = w_j(\boldsymbol{\theta}) - w_j(\boldsymbol{\theta}_0),$$

and $\gamma = \nu(\boldsymbol{\theta}_0) - \nu(\boldsymbol{\theta})$. Then, it is easily seen that $w_j(\boldsymbol{\theta}) = w_j(\boldsymbol{\theta}_0)$, $j = 1, \ldots, p$, and $\nu(\boldsymbol{\theta}) = \nu(\boldsymbol{\theta}_0)$. Moreover, we have either $c_j(\boldsymbol{\theta}) = c_j(\boldsymbol{\theta}_0), \forall j \in \{1, \ldots, p\}$, either $c_j(\boldsymbol{\theta}) = -c_j(\boldsymbol{\theta}_0)$, $\forall j \in \{1, \ldots, p\}$. From the fact that there exists $j_0 \in \{1, \ldots, p\}$ such that the function c_{j_0} is positive and $c_{j_0}(\boldsymbol{\theta}_0) > 0$, we can only have $c_j(\boldsymbol{\theta}) = c_j(\boldsymbol{\theta}_0)$, $j = 1, \ldots, p$.
- Secondly, assume that $\sigma = 0$. By the third point of Lemma 6.3 applied with $\gamma = \nu(\boldsymbol{\theta}_0) - \nu(\boldsymbol{\theta})$ and $\gamma_j = w_j(\boldsymbol{\theta}) - w_j(\boldsymbol{\theta}_0)$, $1 \leq j \leq p$, we get $\nu(\boldsymbol{\theta}) = \nu(\boldsymbol{\theta}_0)$ and $w_j(\boldsymbol{\theta}) = w_j(\boldsymbol{\theta}_0)$, $j = 1, \ldots, p$.

The final conclusion, $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, follows from H5.

Now we can give Theorem 4.1 demonstration. In fact, it is done in a very classical way. By Lemma 6.1, we have:

$$\sup_{\boldsymbol{\theta} \in \Theta} |Q_T(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})| \xrightarrow[T \to \infty]{} 0.$$

Lemma 6.4 can be used to show that

$$Q(\boldsymbol{\theta}_0) < Q(\boldsymbol{\theta}), \quad \forall \boldsymbol{\theta} \in \Theta \setminus \{\boldsymbol{\theta}_0\}$$

(see for example the proof of proposition 2.1 in [15]).
From these last two properties, a classical compactness argument leads to the strong consistency of $\hat{\boldsymbol{\theta}}_T$ (see for example [23, Theorem 2.2.1, p. 19]). ☐

6.6. Proof of Theorem 4.2

Let $t \in \mathbb{Z}$. In Section 4.1, $q_t(\boldsymbol{\theta})$ has been defined as:

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

So the first and second derivatives are:

$$\nabla q_t(\boldsymbol{\theta}) = \frac{\nabla V_t(\boldsymbol{\theta})}{V_t(\boldsymbol{\theta})} \left(1 - \frac{(X_t - m_t(\boldsymbol{\theta}))^2}{V_t(\boldsymbol{\theta})} \right) - 2\frac{(X_t - m_t(\boldsymbol{\theta}))\nabla m_t(\boldsymbol{\theta})}{V_t(\boldsymbol{\theta})} \tag{6.1}$$

$$\nabla^2 q_t(\boldsymbol{\theta}) = \frac{1}{V_t(\boldsymbol{\theta})^2} \left[\nabla V_t(\boldsymbol{\theta}) \nabla V_t(\boldsymbol{\theta})^\top \left(2\frac{(X_t - m_t(\boldsymbol{\theta}))^2}{V_t(\boldsymbol{\theta})} - 1 \right) \right.$$

$$+ \nabla^2 V_t(\boldsymbol{\theta}) \left(1 - \frac{(X_t - m_t(\boldsymbol{\theta}))^2}{V_t(\boldsymbol{\theta})} \right) + 2(X_t - m_t(\boldsymbol{\theta}))\nabla m_t(\boldsymbol{\theta}) \nabla V_t(\boldsymbol{\theta})^\top$$

$$+ 2V_t(\boldsymbol{\theta})\nabla m_t(\boldsymbol{\theta}) \nabla m_t(\boldsymbol{\theta})^\top - 2V_t(\boldsymbol{\theta})(X_t - m_t(\boldsymbol{\theta}))\nabla^2 m_t(\boldsymbol{\theta})$$

$$\left. + 2\left(X_t - m_t(\boldsymbol{\theta}) \right) \nabla V_t(\boldsymbol{\theta}) \nabla m_t(\boldsymbol{\theta})^\top \right] \tag{6.2}$$

Lemmas 6.5 to 6.7 give important properties of $\nabla q_t(\boldsymbol{\theta})$ and $\nabla^2 q_t(\boldsymbol{\theta})$. They are required to prove Theorem 4.2

Lemma 6.5. *For all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*, the sequences* $\{\nabla q_t(\boldsymbol{\theta})\}_t$ *and* $\{\nabla^2 q_t(\boldsymbol{\theta})\}_t$ *are ergodic and stationary.*

Proof. We use the same argument than the one we gave in Lemma 6.1 proof to show that the sequence $(q_t(\boldsymbol{\theta}))_t$ is stationary and ergodic.

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

Lemma 6.6. *We have:*

$$\mathrm{E}\left[\|\nabla q_0(\boldsymbol{\theta}_0)\|^2 \right] < \infty \text{ and } \mathrm{E}\left[\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \nabla^2 q_0(\boldsymbol{\theta}) \right\| \right] < \infty.$$

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

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

Proof of the first assertion: $\mathrm{E}\left[\|\nabla q_0(\boldsymbol{\theta}_0)\|^2 \right] < \infty.$

- We first observe that the ratio:

$$\frac{(X_0 - m_0(\boldsymbol{\theta}_0))\nabla m_0(\boldsymbol{\theta}_0)}{V_0(\boldsymbol{\theta}_0)}$$

is square integrable. Indeed, as

$$\mathrm{E}_{\mathcal{F}_{-1}}\left[\frac{(X_0 - m_0(\boldsymbol{\theta}_0))^2 \|\nabla m_0(\boldsymbol{\theta}_0)\|^2}{V_0(\boldsymbol{\theta}_0)^2} \right] = \frac{\|\nabla m_0(\boldsymbol{\theta}_0)\|^2}{V_0(\boldsymbol{\theta}_0)} \le \frac{\|\nabla m_0(\boldsymbol{\theta}_0)\|^2}{h}$$

there exist positive constants $d_0, \ldots, d_p$ such that

$$\mathrm{E}_{\mathcal{F}_{-1}}\left[\frac{(X_0 - m_0(\boldsymbol{\theta}_0))^2 \, \|\nabla m_0(\boldsymbol{\theta}_0)\|^2}{V_0(\boldsymbol{\theta}_0)^2}\right] \leq d_0 + \sum_{j=1}^{p} d_j X_{-j}^2$$

and the integrability follows from $\mathrm{E}\left[X_0^4\right] < \infty$.

- Now, we show that $\frac{\nabla V_0(\boldsymbol{\theta}_0)}{V_0(\boldsymbol{\theta}_0)}\left(1 - \frac{(X_0 - m_0(\boldsymbol{\theta}_0))^2}{V_0(\boldsymbol{\theta}_0)}\right)$ is square integrable.
  As $V_0(\boldsymbol{\theta}_0) \geq h$ a.s. and $\mathrm{E}\left[X_0^4\right] < \infty$, it is easily seen that the ratio $\frac{\nabla V_0(\boldsymbol{\theta}_0)}{V_0(\boldsymbol{\theta}_0)}$ is square integrable. Then it is enough to show that the variable $V_0(\boldsymbol{\theta}_0)^{-2}\nabla V_0(\boldsymbol{\theta}_0)(X_0 - m_0(\boldsymbol{\theta}_0))^2$ is square integrable. Let

$$C_0(\boldsymbol{\theta}_0) = (X_0 - m_0(\boldsymbol{\theta}_0))^2.$$

As $\nabla V_0(\boldsymbol{\theta}_0)$ is square integrable and measurable with respect to $\mathcal{F}_{-1}$, it is sufficient to show that the random variable $V_0^{-4}(\boldsymbol{\theta}_0)\mathrm{E}_{\mathcal{F}_{-1}}\left[C_0^2(\boldsymbol{\theta}_0)\right]$ is bounded. We notice that:

$$\mathrm{E}_{X_{-1}=x_{-1},\ldots,X_{-p}=x_{-p}}\left[C_0(\boldsymbol{\theta}_0)^2\right]$$

$$= \left\|\sum_{j=1}^{p} b_j(\boldsymbol{\theta}_0) \circ x_{-j} + \xi_0 \sum_{j=1}^{p} c_j(\boldsymbol{\theta}_0) \circ x_{-j} - \sum_{j=1}^{p} w_j(\boldsymbol{\theta}_0)\left|x_{-j}\right| + \eta_t\right\|_4^4$$

$$\leq \left[\sum_{j=1}^{p} \left\|Y_{0,0}^{(j)}\right\|_4 \left|x_{-j}\right| + \|\xi_0\|_4 \left(\sum_{j=1}^{p} \left\|Z_{0,0}^{(j)}\right\|_4 \left|x_{-j}\right|\right) + \sum_{j=1}^{p} w_j(\boldsymbol{\theta}_0)\left|x_{-j}\right| + \|\eta_0\|_4\right]^4.$$

We deduce that there exist constants $d_0, \ldots, d_p$ such that

$$\mathrm{E}_{\mathcal{F}_{-1}}\left[C_0(\boldsymbol{\theta}_0)^2\right] \leq d_0 + \sum_{j=1}^{p} d_j \left|X_{-j}\right|^4.$$

Since $V_0(\boldsymbol{\theta}_0) \geq w_j(\boldsymbol{\theta}_0)\left|X_{-j}\right| \wedge h$ for $j = 1, \ldots, p$, it follows that:

$$V_0(\boldsymbol{\theta}_0)^{-4}\mathrm{E}_{\mathcal{F}_{-1}}\left[C_0(\boldsymbol{\theta}_0)^2\right] \leq \frac{d_0}{h^4} + \sum_{j=1}^{p} \frac{d_j}{w_j(\boldsymbol{\theta}_0)^4}.$$

We have shown that $V_0(\boldsymbol{\theta}_0)^{-4}\mathrm{E}_{\mathcal{F}_{-1}}\left[C_0(\boldsymbol{\theta}_0)^2\right]$ is bounded.

Proof of the second assertion: $\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla^2 q_0(\boldsymbol{\theta})\right\|\right] < \infty$. We start the demonstration by showing that

$$\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\Theta}\left\|\nabla V_0(\boldsymbol{\theta})\right\|^2 V_0(\boldsymbol{\theta})^{-3}[X_0 - m_0(\boldsymbol{\theta})]^2\right] < \infty. \tag{6.3}$$

For this, we have:

$$\mathrm{E}_{\mathcal{F}_{-1}}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} V_0^{-3}(\boldsymbol{\theta})[X_0 - m_0(\boldsymbol{\theta})]^2\right]$$

$$\leq \quad \frac{2}{h\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} V_0^2(\boldsymbol{\theta})}\mathrm{E}_{\mathcal{F}_{-1}}\left[X_0^2 + (\sum_{j=1}^{p}\left\|b_j\right\|_{\infty} X_{-j} + \left\|\mu\right\|_{\infty})^2\right]$$

$$\leq \quad \frac{2\left(m_0(\boldsymbol{\theta}_0)^2 + V_0(\boldsymbol{\theta}_0) + (\sum_{j=1}^{p}\left\|b_j\right\|_{\infty} X_{-j} + \left\|\mu\right\|_{\infty})^2\right)}{h\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} V_0^2(\boldsymbol{\theta})}.$$

Hence, there exist non-negative constants $d, d_1, \ldots, d_p$ such that

$$\mathrm{E}_{\mathcal{F}_{-1}}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} V_0^{-3}(\boldsymbol{\theta})[X_0 - m_0(\boldsymbol{\theta})]^2\right] \leq \frac{1}{\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} V_0^2(\boldsymbol{\theta})}\left(d + \sum_{j=1}^{p} d_j X_{-j}^2\right).$$

As $\dfrac{X_{-j}^2}{\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} V_t^2(\boldsymbol{\theta})} \leq \dfrac{1}{\inf_{j\in\boldsymbol{\Theta}} w_j^2(\boldsymbol{\theta})}$, according to hypothesis H8, we conclude that there exists a constant $M > 0$ such that

$$\mathrm{E}_{\mathcal{F}_{-1}}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} V_0^{-3}(\boldsymbol{\theta})[X_0 - m_0(\boldsymbol{\theta})]^2\right] < M. \tag{6.4}$$

We note that using H8 only for $i = 1, \ldots, d$,

$$\left\|\frac{\partial V_0}{\partial\boldsymbol{\theta}_i}\right\|_{\infty} \leq \quad 2\sigma^2\left(\sum_{j=1}^{p}\left\|c_j\right\|_{\infty}|X_{-j}|\right)\left(\sum_{j=1}^{p}\left\|\frac{\partial c_j}{\partial\boldsymbol{\theta}_i}\right\|_{\infty}|X_{-j}|\right)$$

$$+ \sum_{j=1}^{p}\left\|\frac{\partial w_j}{\partial\boldsymbol{\theta}_i}\right\|_{\infty}|X_{-j}| + \left\|\nu\right\|_{\infty}.$$

Whence, we conclude that if $\mathrm{E}\left[X_0^4\right] < \infty$ then $\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla V_0(\boldsymbol{\theta})\right\|^2\right] < \infty$.

From (6.4) follows (6.3).

We notice that the other terms of (6.2) are uniformly bounded by polynomials of the fourth degree in $|X_{-1}|, \ldots, |X_{-p}|$. This completes the proof.

Lemma 6.7. *The entries of the column vectors of the differential of the function* $\boldsymbol{\theta} \mapsto (m_0(\boldsymbol{\theta}), V_0(\boldsymbol{\theta}))$ *evaluated at* $\boldsymbol{\theta}_0$ *are linearly independent random variables.*

Proof. The proof is done in three steps.

Step 1. Suppose there exist constants $\lambda_1, \ldots, \lambda_d$ such that

$$\sum_{i=1}^{d}\lambda_i\frac{\partial m_0}{\partial\boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) \overset{a.s.}{=} 0 \quad \text{or} \quad \sum_{i=1}^{d}\lambda_i\frac{\partial V_0}{\partial\boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) \overset{a.s.}{=} 0.$$

Since $m_0(\boldsymbol{\theta}_0) = \mu(\boldsymbol{\theta}_0) + \sum_{j=1}^{p} b_j(\boldsymbol{\theta}_0)X_{-j}$, the partial derivatives at $\boldsymbol{\theta}_0$ are:

$$\frac{\partial m_0}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) = \frac{\partial \mu}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) + \sum_{j=1}^{p} \frac{\partial b_j}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0)X_{-j}, \quad i = 1, \ldots, d.$$

Then we have:

$$\sum_{i=1}^{d} \lambda_i \frac{\partial m_0}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) = \sum_{i=1}^{d} \lambda_i \frac{\partial \mu}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) + \sum_{j=1}^{p} \sum_{i=1}^{d} \lambda_i \frac{\partial b_j}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0)X_{-j} \stackrel{a.s.}{=} 0.$$

By the first result of Lemma 6.3, we get:

$$\sum_{i=1}^{d} \lambda_i \frac{\partial \mu}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) = \sum_{i=1}^{d} \lambda_i \frac{\partial b_j}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) \stackrel{a.s.}{=} 0, \quad j = 1, \ldots, p.$$

Step 2. Since $V_0(\boldsymbol{\theta}_0) = \left(\sigma \sum_{j=1}^{p} c_j(\boldsymbol{\theta}_0)X_{-j} \right)^2 + \sum_{j=1}^{p} w_j(\boldsymbol{\theta}_0)|X_{-j}| + \nu(\boldsymbol{\theta}_0)$, the partial derivatives at $\boldsymbol{\theta}_0$ are:

$$\frac{\partial V_0}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) = 2\sigma^2 \left(\sum_{j=1}^{p} c_j(\boldsymbol{\theta}_0)X_{-j} \right) \left(\sum_{j=1}^{p} \frac{\partial b_j}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0)X_{-j} \right)$$
$$+ \sum_{j=1}^{p} \frac{\partial w_j}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0)|X_{-j}| + \frac{\partial \nu}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0),$$

for $i = 1, \ldots, d$. So, we have:

$$\sum_{i=1}^{d} \lambda_i \frac{\partial V_0}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) = 2\sigma^2 \left(\sum_{j=1}^{p} c_j(\boldsymbol{\theta}_0)X_{-j} \right) \left(\sum_{j=1}^{p} \sum_{i=1}^{d} \lambda_i \frac{\partial c_j}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0)X_{-j} \right)$$
$$+ \sum_{j=1}^{p} \sum_{i=1}^{d} \lambda_i \frac{\partial w_j}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0)|X_{-j}| + \sum_{i=1}^{d} \lambda_i \frac{\partial \nu}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) = 0.$$

The last equation can be written as:

$$2\sigma^2 \left(\sum_{j=1}^{p} c_j(\boldsymbol{\theta}_0)X_{-j} \right) \left(\sum_{j=1}^{p} \sum_{i=1}^{d} \lambda_i \frac{\partial c_j}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0)X_{-j} \right)$$
$$+ \sum_{j=1}^{p} \sum_{i=1}^{d} \lambda_i \frac{\partial w_j}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0)|X_{-j}| = - \sum_{i=1}^{d} \lambda_i \frac{\partial \nu}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0).$$

Using the second result of Lemma 6.3 and the assumption that $c_{j_0}(\boldsymbol{\theta}_0) = \beta_{j_0} > 0$, we get

$$2\sigma^2 \sum_{i=1}^{d} \lambda_i \frac{\partial c_j}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) = \sum_{i=1}^{d} \lambda_i \frac{\partial w_j}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) = \sum_{i=1}^{d} \lambda_i \frac{\partial \nu}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_0) = 0, \quad j = 1, \ldots, p.$$

Step 3. By the last two steps and hypothesis H8, we deduce that $\lambda_1 = \cdots = \lambda_d = 0$ which shows that the entries of the vectors of the differential of the bifold function $(m_0(\boldsymbol{\theta}), V_0(\boldsymbol{\theta}))$ evaluated at $\boldsymbol{\theta}_0$ are linearly independent random variables.

We are now ready to prove Theorem 4.2.

Proof of Theorem 4.2. The technique for the proof of this theorem is very classical, we follow the proof given in [23, Theorem 2.2.1, p. 19]. Since $\boldsymbol{\theta} \in \boldsymbol{\Theta}^\circ$, using a Taylor expansion, we get:

$$\mathbf{0} = \nabla Q_T(\hat{\boldsymbol{\theta}}_T) = \nabla Q_T(\boldsymbol{\theta}_0) + \widetilde{\boldsymbol{M}}_T \cdot (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$$

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

$$\widetilde{\boldsymbol{M}}_T(i, j) = \frac{\partial^2 Q_T}{\partial \theta_i \partial \theta_j}(\gamma_i), \quad 1 \leqslant i, j \leqslant d.$$

with $\|\hat{\boldsymbol{\theta}}_T - \gamma_i\| \leqslant \|\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\|$, $\quad i = 1, \ldots, d$. Hence,

$$\sqrt{T} Q_T(\boldsymbol{\theta}_0) = \sqrt{T} \widetilde{\boldsymbol{M}}_T \cdot (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$$

By Lemma 6.5, $\{\nabla^2 q_t(\boldsymbol{\theta}_0)\}_{t \in \mathbb{Z}}$ is an ergodic stationary sequence. By hypothesis H8, its values are in $\mathcal{C}(\boldsymbol{\Theta}, \mathbb{R}^d \times \mathbb{R}^d)$. According to Lemma 6.6, $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\nabla^2 q_0(\boldsymbol{\theta})\|$ is integrable. Then, we can apply Theorem 6.1 and since $\hat{\boldsymbol{\theta}}_T \xrightarrow[T \to \infty]{\text{a.s.}} \boldsymbol{\theta}_0$, we conclude that $\widetilde{\boldsymbol{M}}_T \xrightarrow[T \to \infty]{\text{a.s.}} \boldsymbol{F}_0 = \mathrm{E}[\nabla^2 q_t(\boldsymbol{\theta}_0)]$. More, $\boldsymbol{F}_0$ is non-singular. Indeed:

$$\boldsymbol{F}_0 = \mathrm{E}\left[V_0(\theta_0)^{-2} \left\{\nabla V_0(\theta_0) \nabla V_0(\theta_0)^\top + 2 V_0(\theta_0) \nabla m_0(\theta_0) \nabla m_0(\theta_0)^\top\right\}\right],$$

and by Lemma 6.7, this matrix is positive-definite. More:

$$\sqrt{T} Q_T(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla q_t(\boldsymbol{\theta}_0) \quad \text{and} \quad \mathrm{E}_{\mathcal{F}_{t-1}}[\nabla q_t(\boldsymbol{\theta}_0)] = 0.$$

Since by Lemma 6.6, $\mathrm{E}[\|\nabla q_0(\boldsymbol{\theta}_0)\|^2] < \infty$, the sequence $\{\nabla q_t(\boldsymbol{\theta}_0)\}_t$ is an ergodic stationary $\mathcal{F}_t$-martingale difference sequence of finite variance. Then by [1, Theorem 23.1, p. 206], we have: $\sqrt{T} Q_T(\boldsymbol{\theta}_0) \xrightarrow[T \to \infty]{\mathcal{D}} \mathcal{N}(0, G_0)$, with $G_0 = \mathrm{E}[\nabla q_0(\boldsymbol{\theta}_0) \nabla q_0(\boldsymbol{\theta}_0)']$. Consequently, we get:

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow[T \to \infty]{\mathcal{D}} \mathcal{N}(F_0^{-1} G_0 F_0^{-1}).$$

The expression of G_0 follows from straightforward computations using the expression (6.1). □

Acknowledgements

References

[1] BILLINGSLEY, P. (1999). *Convergence of probability measures* second ed. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York. A Wiley-Interscience Publication.

[2] DION, J.-P., GAUTHIER, G. AND LATOUR, A. (1995). Branching processes with immigration and integer-valued time series. *Serdica Math. J.* **21,** 123–136.

[3] DOUKHAN, P., LATOUR, A. AND ORAICHI, D. (2006). A simple integer-valued bilinear times series models. *Advances in Applied Probability* **38,** 1–20.

[4] DOUKHAN, P. AND TRUQUET, L. (2007). Weakly dependent random fields with infinite memory. *ALEA : Latin American Journal of Probability and Mathematical Statistics* **2,** 111–132.

[5] DROST, F. C., VAN DEN AKKER, R. AND WERKER, B. J. (2007). Note on integer-valued bilinear time series models. *Statistics and Probability Letters*.

[6] DU, J.-G. AND LI, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *J. Times Ser. Anal.* **12,** 129–142.

[7] ENCISO-MORA, V., NEAL, P. AND SUBBA RAO, T. (2006). Efficient order selection algorithms for integer valeud ARMA processes. *Technical report* 2006.415. The University of Manchester School of Mathematics, The University of Manchester, Manchesterm M13 9PL, UK.

[8] FERLAND, R., LATOUR, A. AND ORAICHI, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis* **27,** 923–942.

[9] FOKANIOS, K. (2004). Integer autoregressive and moving average models. *Ann. I.S.U.P.* **48,** 53–58.

[10] FOKANIOS, K. AND FRIED, R. (2008). Intervention effects in INGARCH processes. Working paper.

[11] FREELAND, R. AND MCCAB, B. (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting* **20,** 427– 434.

[12] GATHER, U., SCHETTLINGER, K. AND FRIED, R. (2006). Online signal extraction by robust linear regression. *Computational Statistics* **21,** 33–51.

[13] HEINEN, A. (2001). Modeling Time Series Count Data: The Autoregressive Conditional Poisson Model. *PhD thesis*. Department of Economics, University of California San Diego, California.

[14] HELLSTRÖM, J. (2001). Unit root testing in integer-valued ar(1) models. *Economics Letters* **70,** 9–14.

[15] JEANTHEAU, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* **14,** 70–86.

[16] KEDEM, B. AND FOKANIOS, K. (2002). Regression models for binary time series. In *Modeling uncertainty*. vol. 46 of *Internat. Ser. Oper. Res. Management Sci.* Kluwer Academic Publisher, Boston, MA. pp. 185–199.

[17] KEDEM, B. AND FOKANIOS, K. (2002). *Regression Models for Time Series Analysis*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, New Jersey.

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

[19] MACDONALD, I. L. AND ZUCCHINI, W. (1997). *Hidden Markov and other models for discrete-valued time series* vol. 70 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

[20] MCCAB, B. AND MARTIN, G. (2005). Bayesian predictions of low count time series. *International Journal of Forecasting* **22,** 315–330.

[21] NEAL, P. AND SUBBA RAO, T. (2007). Mcmc for integer-valued ARMA processes. *Journal of Time Series Analysis* **28,** 92–110.

[22] STEUTEL, F. AND VAN HARN, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Prob.* **7,** 893–899.

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