Center for Quantitative Economics

Strict stationarity of Poisson integer-valued ARCH processes of order infinity

Mawuli Segnon [†]

102/2022

[†] Department of Economics, University of Münster, Germany

Strict stationarity of Poisson integer-valued ARCH processes of order infinity

Mawuli Segnon*

Center for Quantitative Economics, University of Münster, Germany

Abstract: This paper establishes necessary and sufficient conditions for the existence of a unique strictly stationary and ergodic solution for integer-valued autoregressive conditional heteroscedasticity (INARCH) processes. We also provide conditions that guarantee existence of higher order moments. The results apply to integer-valued GARCH model, and its long-memory versions with hyperbolically decaying coefficients and turn out to be instrumental on deriving large sample properties of the maximum likelihood estimators of the model parameters.

KEY WORDS: INARCH processes; Stationarity; Ergodicity; Lyapunov exponent; Maximum likelihood estimation

JEL classification: C1, C4, C5

*Corresponding author: Mawuli Segnon, Center for Quantitative Economics, Am Stadtgraben 9, 48143 Münster, Germany. E-mail: segnon@uni-muenster.de.

1 Introduction

In many real-world situations, we have to deal with non-negative integer-valued time series. Such time series are often produced in fields that include economics, insurance, medicine, epidemiology, queueing systems, communications, and meteorology and so on. Examples for the wide range of practical applications are the daily or monthly number of cases in epidemiology, the number of stock market transactions or stock price changes per minute in finance and the number of photon arrivals per microsecond measured in a biological experiment. Their analysis may present some difficulties, however, and if the analysis is based on stochastic models, these models have to reflect the integer peculiarity of the observed series. Various models have been suggested in the literature to tackle the problem of integer-valued time series analysis. These models include the traditional generalized linear model methodology and the state-of-the-art integer-valued autoregressive moving average (INARMA), and integer-valued generalized autoregressive conditional heteroscedasticity (INGARCH) processes. The first modeling approach is very simple and consists of choosing a suitable distribution for count data and an appropriate link function, (see Kedem and Fokianos, 2002). The second group of models are adaptation of the well-known ARMA and GARCH processes in the modeling of continuous-state and discrete-time series to count settings by means of thinning operators (see Weiß, 2008, for a recent review of the thinning operators). These processes are developed to model stationary count data. Therefore, considerable effort has been devoted to provide and prove general conditions that ensure existence and uniqueness of second-order stationary solutions using Hilbert space techniques (see Ferland et al., 2006; Latour, 1998; Doukhan and Wintenberger, 2008; Doukhan et al., 2012; Neumann, 2011). Recently, Sim et al. (2021) provide conditions for ergodicity and consistency of the maximum likelihood estimator for general-order observation-driven models (ODMs). However, recent empirical observations indicate that some important count data in modeling are strictly stationary, and non square-integrable (see Segnon and Stapper, 2019).

The objective of this paper is to establish conditions for strict stationarity and ergodicity of the INARCH processes and existence of higher order moments. These statistical properties are crucial for deriving large sample properties of the maximum likelihood estimators of the model parameters. We make use of the multiplicative ergodic theorem developed by Ruelle (1982) for bounded operators in Hilbert space and show that the necessary and sufficient conditions for stationarity is the negativity of the Lyapunov exponent associated with these processes. Our result applies to the INGARCH model in Ferland et al. (2006), and INFIGARCH and INHYGARCH models in Segnon and Stapper (2019). Since the seminal paper by Bougerol and Picard (1992) the use of the multiplicative ergodic theorem to study the stationarity of ARCH-type processes has become very popular, see Kazakevicius and Leipus (2002); Zerner (2018).

The rest of the paper is organized as follows. Section 2 describes the modeling framework. The main results are provided in Section 3. The conditional maximum likelihood estimator and its asymptotic distribution are presented in Section 4. Section 5 presents the proofs to the main results. Finally, Section 6 concludes.

2 Poisson INARCH(∞) Processes

2.1 Definition

A sequence of integer-valued random variables $\{Y_t\}_{t \in Z}$ is said to be an INARCH(∞) process if:

(i) the distribution of Y_t conditional on the $\sigma-$field $\Omega_{t-1} = \sigma(Y_l, l \leq t-1)$ is Poisson with mean λ_t,

(ii) there exist nonnegative constants c, ψ_i, $1 \leq i \leq \infty$, such that

$$\lambda_t = c + \psi(\mathrm{L})\, Y_t, \tag{1}$$

where $\Pr(\lambda_t > 0) = 1$ and $\psi(\mathrm{L}) = \sum_{i=1}^{\infty} \psi_i \mathrm{L}^i$.

This class of models also includes:

(a) The integer-valued HYGARCH(p, d, q) model for, c is an appropriately defined constant, and

$$\begin{aligned}\psi(\mathrm{L}) &= \left[1 - \frac{\Phi(\mathrm{L})\left(1 + \eta[(1-\mathrm{L})^d - 1]\right)}{\mathbf{B}(\mathrm{L})}\right] \\ &= \sum_{i=1}^{\infty} \psi_i \mathrm{L}^i,\end{aligned} \tag{2}$$

with $\beta_0 > 0$ and $\phi_1, \ldots, \phi_{m-1} \geq 0, \beta_1, \ldots, \beta_q \geq 0$, and $\psi_i \geq 0$ for all i. In Eq. (2), L denotes the lag operator. The lag polynomials are defined as $\Phi(\mathrm{L}) = [1 - \beta(\mathrm{L}) - \alpha(\mathrm{L})] = \sum_{i=1}^{m-1} \phi_i \mathrm{L}^i$, where $m = \max(p, q)$, $\alpha(\mathrm{L}) = \sum_{i=1}^{p} \alpha_i \mathrm{L}^i$, $\beta(\mathrm{L}) = \sum_{j=1}^{q} \beta_j \mathrm{L}^j$ and $\mathbf{B}(\mathrm{L}) = [1 - \beta(\mathrm{L})]$. $\eta \geq 0$ is an amplitude parameter, $d \in [0, 1]$ and $(1-\mathrm{L})^d$ is the fractional differencing operator given by

$$(1-\mathrm{L})^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)\mathrm{L}^k}{\Gamma(-d)\Gamma(k+1)}, \tag{3}$$

where $\Gamma(\cdot)$ is the gamma function.

(b) The integer-valued FIGARCH(p, d, q) model for $\eta = 1$ in Eq. 2.

(c) The integer-valued GARCH(p, q) model for $\eta = 0$ in Eq. 2.

Remark 1. *Segnon and Stapper (2019) show that for $\eta \in (0, 1)$ implies that $\psi(1) < 1$, and thus, the INHYGARCH process is covariance stationary.*

Remark 2. *Ferland et al. (2006) show that the INGARCH(p,q) process exists and is strictly stationary with finite first and second order moments, if and only if the following restriction is met: $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$, which is equivalent to $\sum_{i=1}^{\infty} \psi_i < 1$. In the simple INGARCH$(1, 1)$, $\psi_i = \alpha_1 \beta_1^{i-1}$ for $i \geq 1$ and the stationarity condition is well known to be $\alpha_1 + \beta_1 < 1$, which is* *equivalent to* $\sum_{i=1}^{\infty} \alpha_1 \beta_1^{i-1} < 1$ *in the INARCH representation above. The INGARCH*$(1,1)$ *reduces to an integrated INGARCH*$(1,1)$ *when the sum of the lag coefficients is unity* $(\alpha_1+\beta_1 = 1)$*. Segnon and Stapper (2019) point out that in the INFIGARCH(p,q)* $\sum_{i=1}^{\infty} \psi_i = 1$*. Thus, the process is not covariance stationary. We note that the coefficient* ψ_i *in the INHYGARCH can be approximated by* ci^{-1-d}*, with c appropriately defined.*

Figure 1 illustrates the capacity of the INHYGARCH(1,d,1) model to reproduce various degree of dependence for $\eta = 1$ (INFIGARCH(1,d,1)), $\eta = 0$ (INGARCH(1,1)) and $\eta = 0.8$ (INHYGARCH(1,d,1)). We see that with different values for the amplitude parameter, η, various long range dependencies observed in empirical data can be captured. Furthermore, Table 1 shows that the INHYGARCH(1,d,1) model can reproduce over-dispersion and asymmetry observed in real world data.

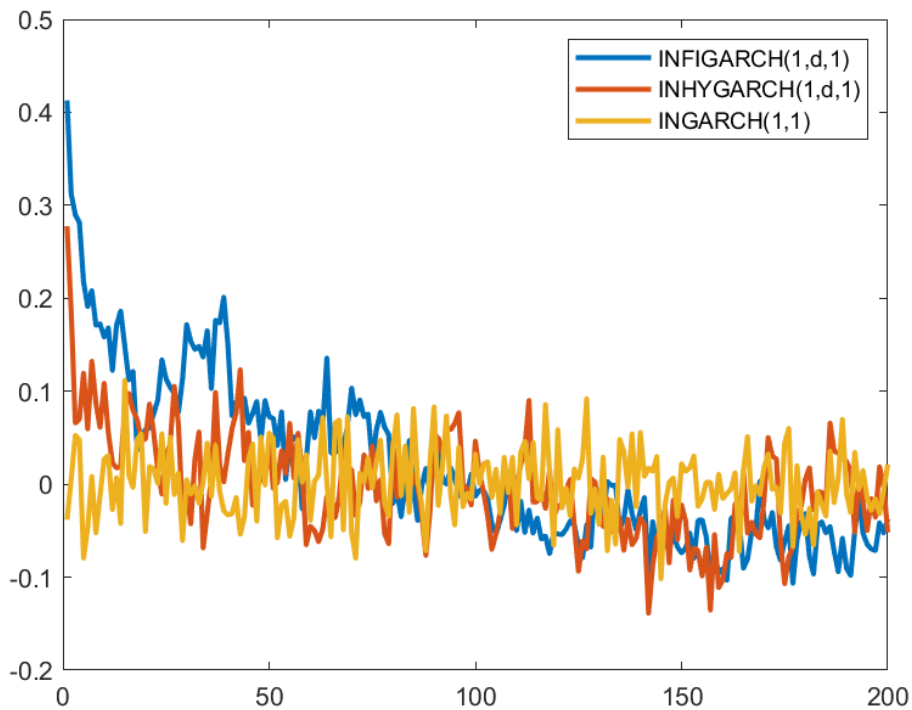

Figure 1: Theoretical ACF for different parameter constellations with baseline setup: $\beta_0 = 2$, $\alpha_1 = 0.3$, $d = 0.4$, $\beta_1 = 0.2$, $\eta = 0.8$ and $n = 500$.

Since the INFIGARCH(p,d,q) process is not covariance stationary, its appears that the INFIGARCH(p,d,q) is not a long memory model in the common sense. However, we aim to show in the next Section that the INARCH representation of the INHYGARCH(p,d,q) and INFIGARCH(p,d,q) processes are strictly stationary and ergodic using a multiplicative ergodic theorem and a Lyapunov exponent. Towards this end, we first look at the construction of an INARCH process.

Table 1: Descriptive statistics of simulated data

	INFIGARCH	INHYGARCH	INGARCH
Overdispersion	11.812	3.736	1.522
Skewness	0.210	0.473	0.663
Kurtosis	3.028	3.301	3.452

Note: The statistics reported in the Table are the averages. The results are based on 100 replications of simulated data with size (n=500) with the following parameters: $\beta_0 = 2$, $\alpha_1 = 0.3$, $d = 0.4$, $\beta_1 = 0.2$, $\eta = 0.8$.

2.2 Construction

Let $\{u_t\}_{t \in Z}$ be a sequence of independent random variables with values in N (N is the set of non-negative integers) with common mean ω. For each $t \in Z$ and $i \in N$, let $\xi_t^{(i)} = \{\xi_{t,j}^{(i)}\}_{j \in N}$ represent a sequence of independent random variables having a common mean ψ_i. All the variables u_s, $\xi_{t,j}^{(i)}$, $(s \in Z, t \in Z, i \in N$ and $j \in N)$ are assumed to be mutually independent. Using these random variables, we introduce a sequence of random variables $\{Y_t^{(n)}\}$ that may be considered as successive approximations of Y_t:

$$Y_t^{(n)} = \begin{cases} 0, & \text{if } n < 0; \\ u_t, & \text{if } n = 0; \\ u_t + \sum_{i=1}^{n} \sum_{j=1}^{Y_{t-i}^{(n-i)}} \xi_{t-i,j}^{(i)} & \text{if } n > 0. \end{cases} \tag{4}$$

From (4) we can see that $Y_t^{(n)}$ is a finite sum of independent Poisson variables. So, the expectation and the variance of $Y_t^{(n)}$ are well defined. In the next Section we want to show that $Y_t^{(n)}$, as $n \to \infty$, admits an almost sure limit Y_t and that the limiting process $\{Y_t\}_{t \in Z}$ satisfies (1), see Proposition 1. Given this result, we want to show that under mild conditions the approximated process is strictly stationary, ergodic and has moments of any order. Then, these statistical properties for the original process is obtained by a limiting argument (Proposition 1) connecting the two representations.

3 Stationarity of INARCH(∞) Processes

To prove the strict stationarity of $\{Y_t\}$ we first show that for any fixed n, $Y_t^{(n)}$ is strictly stationary.

3.1 Some basic definitions and results

Definition 1. *Let $\{z_j\}_{j \in N}$ be a sequence of independent and identically distributed non-negative integer-valued random variables with mean ψ and finite variance σ^2 which is independent of a non-negative integer-valued random variable y. The generalized Steutel and van Harn operator, $\psi \diamond$, is defined as*

$$\psi \diamond y = \begin{cases} \sum_{i=1}^{y} z_i & \text{if } y > 0; \\ 0 & \text{if } y = 0. \end{cases} \quad (5)$$

Remark 3. *The sequence $\{z_j\}_{j\in N}$ is called a counting sequence. Let $\alpha\diamond$ be another operator based on a counting sequence $\{x_j\}_{j\in N}$. Both operators $\psi\diamond$ and $\alpha\diamond$ are said to be independent if and only if the counting sequences $\{z_j\}_{j\in N}$ and $\{x_j\}_{j\in N}$ are mutually independent.*

Using the operator from Eq. 5, we may rewrite the sequence of random variables $\{Y_t^{(n)}\}_{n\in N}$ as

$$Y_t^{(n)} = \sum_{i=1}^{n} E\left(\xi_{t-i}^{(i)}\right) \diamond Y_{t-i}^{(n-i)} + u_t, \quad n > 0, \quad (6)$$

where $E\left(\xi_{t-i}^{(i)}\right) = \psi_i$.

Proposition 1. *If $\psi(1) < 1$ then the sequence $\{Y_t^{(n)}\}_{n\in N}$ has an almost sure limit.*

Proposition 2. *Let $\mathbf{C}_t = \{c_{i,j}\}_{1<i,j<n}$ be an finite-dimensional random matrix given by*

$$\mathbf{C}_t = \begin{pmatrix} \xi_{t-1}^{(1)} & \xi_{t-2}^{(2)} & \dots & \xi_{t-n}^{(n)} \\ 1 & 0 & \dots & \dots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \dots & 1 & 0 \end{pmatrix}, \quad (7)$$

$\mathbf{Z}_{t+1} = \left(Y_t^{(n)}, Y_{t-1}^{(n-1)}, Y_{t-2}^{(n-2)}, \dots Y_{t-n}^{(0)}\right)'$ *and* $\mathbf{U}_{t+1} = (u_t, 0, \dots)'$.
Then, (6) has a stationary and ergodic solution if and only if

$$\mathbf{Z}_{t+1} = E(\mathbf{C}_{t+1}) \diamond \mathbf{Z}_t + \mathbf{U}_{t+1}, \quad t \in Z, \quad (8)$$

has a stationary and ergodic solution where

$$E(\mathbf{C}_t) \diamond := \tilde{c}_{ij} \diamond = \begin{cases} \psi\diamond & i = 1; \\ 1\diamond & i = j+1 \\ 0\diamond & \text{otherwise}. \end{cases}$$

Lemma 1. *Let $\psi(z) = z^n - \alpha_1 z^{n-1} - \dots - \alpha_{n-1} z - \alpha_n$ with $\sum_{k=1}^{n} |\alpha_k| \le 1$ and $\alpha_n > 0$. Then the roots of $\psi(z)$ are all inside the unit circle.*

Lemma 2. *(Lemma 2.1 in Bougerol and Picard (1992)) Let $\{\mathbf{A}_n, n \in Z\}$ be a sequence of independent, identically distributed, random matrices such that $E(\log^+ \|\mathbf{A}_0\|)$ is finite. If, almost surely,*

$$\lim_{n\to+\infty} \frac{1}{n} \log \|\mathbf{A}_0 \mathbf{A}_{-1} \cdots \mathbf{A}_{-n}\| = 0,$$

then the top Lyapunov exponent associated with this sequence is strictly negative.

Proposition 3. *The process defined in Eq. (8) has a unique strictly stationary and ergodic solution if and only if the top Lyapunov exponent γ associated with the random matrices $\{\mathbf{C}_t\}_{t\in Z}$ is strictly negative. The unique strictly stationary solution $(\mathbf{Z}_t)_{t\in Z}$ of (8) is given by*

$$\mathbf{Z}_t = \sum_{k=0}^{\infty} E\left(\mathbf{C}_t \mathbf{C}_{t-1} \ldots \mathbf{C}_{t-k+1}\right) \diamond \mathbf{U}_{t-k}. \tag{9}$$

Corollary 1. *The process $\{Y_t\}_{t \in Z}$ is a strictly stationary and ergodic process.*

Proposition 3 holds its validity for any fixed number n, see Bougerol and Picard (1992) for GARCH(p,q) models. When $n \to \infty$, then $\mathbf{C}_t$ becomes an infinite dimensional random matrix and as pointed out by Schaumlöffel (1991) the multiplicative ergodic theorem of Oseledec (1968) cannot easily be extended to an infinite-dimensional context without any additional assumptions. The reason is that in infinite dimensions the orbits of a linear operator can be quite complicated. For that reason we need here additional assumptions to guarantee the validity of the Proposition 3. Following Ruelle (1982) we show that the compactness of the linear operator $\mathbf{C}$ is the necessary and sufficient condition for the strict negativity of the associated top Lyapunov exponent.

Let H denotes a separable infinite-dimensional Hilbert space and $B(H)$ the algebra of all bounded operators on H. The space $H^{\wedge q}$ is the q^{th} exterior power of H and it consists of the completely antisymmetric elements of the Hilbert space tensor product of q copies of H. Let $\{e_n\}$ be any orthonormal basis $\mathcal{B}$ for H. According to Definition 7 in Van Barel et al. (1999) [Definition of extended infinite companion matrix], when $n \to \infty$, the $\infty \otimes \infty$ matrix $\mathbf{C} = (c_{ij})$, $c_{ij} = (\mathbf{C}e_j|e_i)$, $i, j \in N$, see Eq. 7, has a block structure that corresponds to the block structure of the basis $\mathcal{B}$. Formally, we have

$$\mathbf{C} = \left[\mathbf{C}_{i,j}\right]_{i,j=1}^{\infty}, \tag{10}$$

where the blocks $C_{i,j}$ are square of order K.

To prove the compactness of $\mathbf{C}$ we need first to prove that the linear bounded operator $\mathbf{C}$ has a tri-block diagonal matrix representation with finite blocks. This idea has been put forward by Bakic and Guljas (1999) and the following Proposition shows that a tri-block diagonal matrix representation for the bounded operator $\mathbf{C}$ can be obtained from any orthonormal basis of the Hilbert space H by an arbitrary small Hilbert-Schmidt perturbation.

Proposition 4. *Suppose $\mathbf{C} \in B(H)$ be an operator having a tri-block diagonal matrix*

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{1,1} & \mathbf{C}_{1,2} & 0 & & & & \\ \mathbf{C}_{2,1} & \mathbf{C}_{2,2} & \mathbf{C}_{2,3} & 0 & & & \\ 0 & \mathbf{C}_{3,2} & \ddots & \ddots & \ddots & & \\ & 0 & \ddots & \mathbf{C}_{n,n} & \mathbf{C}_{n,n+1} & 0 & \\ & & \ddots & \mathbf{C}_{n+1,n} & \mathbf{C}_{n+1,n+1} & \ddots & \ddots \\ & & & 0 & \ddots & \ddots & \\ & & & & \ddots & & \end{pmatrix}, \tag{11}$$

according to the decomposition $H = \oplus_{n=1}^{\infty} H_n$ onto finite square dimensional subspaces H_n and

$$\lim_n \|\mathbf{C}_{n,n}\| = \lim_n \|\mathbf{C}_{n+1,n}\| = \lim_n \|\mathbf{C}_{n,n+1}\| = 0. \tag{12}$$

Then $\mathbf{C}$ *is a compact operator.*

The following assumptions are provided in Ruelle (1982). Let (M, Ω, p) be a probability space and $\vartheta : M \to M$ a measurable $p-$preserving transformation on H. Let $\mathbf{C} : \Omega \to L(H)$ be measurable to the bounded operators such that

(1.1) $\log^+ \|\mathbf{C}(\cdot)\| \in L^1(M, p)$.

Let us define

$$\mathbf{C}_x^t = \mathbf{C}(\vartheta^{t-1}x)\cdots\mathbf{C}(\vartheta x)\mathbf{C}(x).$$

Then, there exists a subset $\Gamma^+ \subset M$ such that $\vartheta\Gamma^+ \subset \Gamma^+$, $p(\Gamma^+) = 1$ and

(1.2) $\limsup_{t\to\infty} \log \|\mathbf{C}(\vartheta^{t-1}x)\| \leq 0$, if $x \in \Gamma^+$.

(1.3) Furthermore, there exist $\vartheta-$invariant functions $l_q^+ : \Gamma^+ \to R \bigcup\{-\infty\}$ such that

$$\lim \frac{1}{t} \log \|(\mathbf{C}_x^t)^{\wedge q}\| = l_q^+$$

if $x \in \Gamma^+$, for all integers $q > 0$.

Proposition 5. *Suppose that the assumptions (1.1), (1.2) and (1.3) hold and that* $\mathbf{C}$ *is compact. Then the top Lyapunov exponent* γ *associated with the infinite random matrices* $\{\mathbf{C}_t\}_{t\in Z}$ *is strictly negative* $(-\infty)$ *and the process defined in Eq. (8) has a unique strictly stationary and ergodic solution,* $(\mathbf{Z}_t)_{t\in Z}$*, that is given by*

$$\mathbf{Z}_t = \sum_{k=0}^{\infty} E\left(\mathbf{C}_t\mathbf{C}_{t-1}\ldots\mathbf{C}_{t-k+1}\right) \diamond \mathbf{U}_{t-k}. \tag{13}$$

Corollary 2. *Let assume that the support of* f *is unbounded,* $f(\{0\}) = 0$ *and all the coefficients* ψ *are nonnegative. Then, if* $\sum_{i=1}^n \psi_i = 1$*, then the INARCH process defined in Eq. (8) has a unique stationary solution.*

3.2 Moments of the INARCH Processes

It is crucial for statistical inference to know whether the unique stationary solution has moments of higher order. To derive conditions for the existence of higher order moments for INARCH models we use the state space representation of the successive approximated process in (8). The following proposition guarantees the existence of higher order moments.

Proposition 6. *Let* $m \in N^*$. *Then the mth moment of* $Y_t^{(n)}$ *is finite if and only if the spectral radius of the matrix* $E\left(\mathbf{C}_t^{\otimes m}\right)$ *is strictly less than 1, where* $\mathbf{C}^{\otimes m} = \mathbf{C}\otimes\mathbf{C}\otimes\cdots\otimes\mathbf{C}$ *(m factors),* $\rho(\mathbf{C}) = \min\{|$*eigenvalues of a matrix* $\mathbf{C})\}$ *and* $\mathbf{C}_t$ *is defined by Eq. (7).*

4 Likelihood Inference

Fokianos et al. (2009) and Fokianos and Tjøstheim (2012) have studied in details the principles of likelihood inference in the linear and nonlinear models. Denoting $\boldsymbol{\theta}$ =

$(\omega, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q, d, \eta)'$ the vector of unknown parameters in the INHYGARCH model, assume that $\boldsymbol{\theta}_0$ is the true parameter vector and is in the interior of the compact set Θ. Then the conditional likelihood function for $\boldsymbol{\theta}$, given the starting value λ_0 in terms of observations $y_{-\infty}, ..., y_0, y_1, ..., y_n$, is given by

$$\mathrm{L}(\boldsymbol{\theta}) = \prod_{t=1}^{n} \frac{\lambda_t^{Y_t}(\boldsymbol{\theta})}{Y_t!} \exp(-\lambda_t(\boldsymbol{\theta})) \,, \tag{14}$$

where $\lambda_t(\boldsymbol{\theta}) = \omega + \psi^{\mathrm{INHY}}(\mathrm{B})Y_t$ with $\lambda_t = \lambda_t(\boldsymbol{\theta}_0)$. Thus, the log-likelihood function is given, up to a constant, by

$$l\,(\boldsymbol{\theta}) = \sum_{t=1}^{n} l_t\,(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(Y_t \log \lambda_t(\boldsymbol{\theta}) - \lambda_t(\boldsymbol{\theta})\right), \tag{15}$$

and the score function is defined by

$$\begin{aligned} S_n\,(\boldsymbol{\theta}) &= \frac{\partial l\,(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{\partial l_t\,(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \sum_{t=1}^{n} \left(\frac{Y_t}{\lambda_t(\boldsymbol{\theta})} - 1 \right) \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \end{aligned} \tag{16}$$

The conditional maximum likelihood estimates are found by maximizing the log likelihood $l(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}}_{\mathrm{CML}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\; l(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \sum_{t=1}^{n} l_t(\boldsymbol{\theta}) \,. \tag{17}$$

Furthermore, the Hessian matrix for our proposed model is obtained by further differentiation of the score equations (16). That is

$$\begin{aligned} H_n\,(\boldsymbol{\theta}) &= -\sum_{t=1}^{n} \frac{\partial^2 l_t\,(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \\ &= \sum_{t=1}^{n} \frac{Y_t}{\lambda_t^2(\boldsymbol{\theta})} \left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' \\ &\quad - \sum_{t=1}^{n} \left(\frac{Y_t}{\lambda_t(\boldsymbol{\theta})} - 1 \right) \frac{\partial \lambda_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}. \end{aligned} \tag{18}$$

Let denote $D_t = \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, $P_t(\boldsymbol{\theta}) = -\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$, $\Sigma = \mathrm{E}\,[P_t(\boldsymbol{\theta}_0)]$, $\Omega = \mathrm{E}\left[D_t(\boldsymbol{\theta}_0) D_t'(\boldsymbol{\theta}_0)\right]$ and $V_0(\varsigma) = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \varsigma\}$. We make the following assumptions:

Assumption 1. *(Ling and McAleer (2010))*

(i) $E \sup_{\boldsymbol{\theta}\in\Theta} [l_t(\boldsymbol{\theta})] < \infty$*, and* $E[l_t(\boldsymbol{\theta})]$ *has a unique maximizer at* $\boldsymbol{\theta}_0$*;*

(ii) $D_t(\boldsymbol{\theta}_0)$ *is a martingale difference in terms of* $\mathcal{F}_t$ *with* $0 < \Omega < \infty$*;*

(iii) $\Sigma > 0$ *and* $E \sup_{\boldsymbol{\theta} \in V_0(\varsigma)} \|P_t(\boldsymbol{\theta})\| < \infty$ *for some* $\varsigma > 0$.

Assumption 2. *(Ling and McAleer (2010), for long-memory time series).*
For some $v > 0$*, it follows that*

(i) $E \sup_{\Theta} \|l_t(\boldsymbol{\theta}) - \tilde{l}_t(\boldsymbol{\theta})\| = O\left(\frac{1}{t^v}\right)$;

(ii) $\lim_{l \to \infty} \mathrm{P}\left(\max_{l \le n < \infty} \frac{1}{\sqrt{n}} \| \sum_{t=1}^{n} \left[D_t(\boldsymbol{\theta}_0) - \tilde{D}_t(\boldsymbol{\theta}_0)\right] \| > \epsilon\right) = 0$*, for any* $\epsilon > 0$ *and*

(iii) $E \sup_{\Theta} \|P_t(\boldsymbol{\theta}) - \tilde{P}_t(\boldsymbol{\theta})\| = O\left(\frac{1}{t^v}\right)$.

We note that the dimension of the initial value Y_0 is infinite in our modeling framework. So, we replace it with some constant $\tilde{Y}_0$ as suggested in Ling and McAleer (2010) and we denote $l_t(\boldsymbol{\theta})$, $D_t(\boldsymbol{\theta}_0)$ and $P_t(\boldsymbol{\theta})$ with the initial value $\tilde{Y}_0$ by $\tilde{l}_t(\boldsymbol{\theta})$, $\tilde{D}_t(\boldsymbol{\theta}_0)$ and $\tilde{P}_t(\boldsymbol{\theta})$, respectively.

Assumption 3. *The long memory parameter* $d \in (0,1)$*,* $\eta \in (0,1)$*, all roots of* $\Phi(B)$ *and* $\mathbf{B}(B)$ *are outside the unit circle,* $\phi_m \neq 0$*,* $\beta_q \neq 0$ *with* $m = \max(p,q)$ *and* $\Phi(B)$ *and* $\beta(B)$ *have no common root.*

Proposition 7. *Under Assumptions (1), (2) and (3), we have*

1.
$$\hat{\boldsymbol{\theta}}_{CML} \to \boldsymbol{\theta}_0 \ a.s.$$

2.
$$\hat{\boldsymbol{\theta}}_{CML} = \boldsymbol{\theta}_0 + O\left[\left(\frac{\log \log n}{n}\right)^{1/2}\right]$$

3.
$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{CML} - \boldsymbol{\theta}_0\right) \to_L N\left(0, \Sigma^{-1}\Omega\Sigma^{-1}\right).$$

4.1 Monte Carlo Simulation

We investigate the performance of the CML estimation for INFIGARCH and INHYGARCH models in a simulation study. The most interesting parameter is d, since it controls long-memory behavior. Parameters less interesting for interpretation are chosen to be the same for all models considered, namely $\alpha_1 = 0.2$, $\beta_1 = 0.5$ and $\eta = 0.85$. Since both models aim at use for high frequency financial data, there is no problem collecting enough data. Therefore, we consider the sample sizes $n \in \{10000, 20000, 50000\}$, simulated with a burn-in of length 10000.

The intercept is chosen such that the marginal mean is between 15 and 20, for better comparability among settings. For each setting, 1000 time series are simulated and the parameters estimated. A comprehensive summary of the simulation is displayed in Table 2 in the appendix. As expected, the estimates for a larger number of observations, have less bias and variation. We also observe a decrease in the root mean-squared errors (RMSE) as the sample size increases (we do not report the results, but they are available upon requests).

5 Proofs

Proof of Proposition 1. We closely follow the Proof of Proposition 2 in Ferland et al. (2006), Page 928. It follows from Eq. (4) that $Y_t^{(n)}$ is obtained through a cascade of thinning operations along the sequence $\{u_t\}_{t\in Z}$. So, the expectation and the variance of $Y_t^{(n)}$ are well defined and given by

$$\begin{aligned}\mu_n &= E\left(u_t + \sum_{i=1}^{n}\sum_{j=1}^{Y_{t-i}^{(n-i)}} \xi_{t-i,j}^{(i)}\right) \\ &= \omega + \sum_{i=1}^{n} E\left(\sum_{j=1}^{Y_{t-i}^{(n-i)}} \xi_{t-i,j}^{(i)}\right)\end{aligned} \tag{19}$$

Let (Ω, F, P) be the common probability space on which the relevant random variables are defined. Because $Y_j^{(n)}$ is a non-decreasing sequence of non-negative integers, we have

$$\forall \omega \in \Omega, \lim_{n\to\infty} Y_t^{(n)}(\omega) = Y_t \tag{20}$$

which is either finite or infinite. We will show that the set

$$A_\infty = \{\omega : Y_t(\omega) = \infty\} = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k = \limsup_n A_n \tag{21}$$

is of probability zero, where

$$A_n = \left\{\omega : Y_t^{(n)}(\omega) - Y_t^{(n-1)}(\omega) > 0\right\}, \text{ for } n > 1. \tag{22}$$

On the one hand:

$$E\left(Y_t^{(n)} - Y_t^{(n-1)}\right) \geq \sum_{k=1}^{\infty} \Pr\left\{\omega : Y_t^{(n)}(\omega) - Y_t^{(n-1)}(\omega) = k\right\} = \Pr(A_n). \tag{23}$$

On the other hand:

$$E\left(Y_t^{(n)} - Y_t^{(n-1)}\right) = \mu_n - \mu_{n-1} \equiv \nu_n. \tag{24}$$

Obviously, the sequence $\{\nu_n\}$ satisfies a homogeneous finite difference equation with a characteristic polynomial, namely $\psi(z)$, that has all its roots outside the unit circle. As shown in Brockwell and Davis (1991), Section 3.6, sequence $\{\nu_n\}$ tends towards zero with a geometric rate as $n \to \infty$. In other words, a constant $Q \geq 0$ and a constant $0 < \alpha < 1$ exist such that $\nu_n \leq Q\alpha^n$. Since $\Pr\{A_n\} \leq \nu_n$ we get

$$\sum_{n=1}^{\infty} \Pr\{A_n\} \leq Q \sum_{n=1}^{\infty} \alpha^n < \infty. \tag{25}$$

By the Borel-Cantelli lemma, $\Pr\{A_\infty\} = 0$. ■

Proof of Proposition 2. Eq. (8) is a state-space representation of (6), and thus, any stationary solution of (8) is also a stationary solution of (6) and vice versa. Analogously, any ergodic solution of (8) is also an ergodic solution of (6), and vice versa. The proof of the ergodicity follows from Lemma A 1.2.7 in Brandt et al. (1990). ■

Proof of Lemma 1. Let us consider the unit circle $\zeta = \{z : |z| = 1\}$ and suppose $\sum_{k=1}^{n} |\alpha_k| < 1$. The functions $h(z) = z^n$ and $T(z) = -(\alpha_1 z^{n-1} + \alpha_2 z^{n-2} + \cdots + \alpha_n)$ are both analytic inside and on ζ. Hence, on ζ,

$$|T| \leq \sum_{k=0}^{r-1} |\alpha_{n-k} z^k| \leq \sum_{k=0}^{n-1} |\alpha_{n-k}| < 1 = |h|.$$

Based on the theorem of Rouché, $h(z)$ and $h(z) + T(z)$ have the same number of zeros inside ζ. But h has n zeros inside ζ. Therefore, we conclude that all roots of $\alpha(z)$ are inside the unit circle. ■

Proof of Proposition 3. Suppose that the top Lyapunov exponent γ is strictly negative. We can see that the random matrices $\{\mathbf{C}_t\}$ in Eq. 7 consist of independent and identically distributed non-negative integer-valued random variables, $\xi_t^{(i)}$, with the baseline distribution f (Poisson) and with a finite mean, ψ_i and variance. This means that all the coefficients of these matrices are integrable. Furthermore, the random vectors $\{\mathbf{U}_t\}_{t\in Z}$ contain *i.i.d.* non-negative integer-valued random variables and therefore are also integrable. All these imply that $E(\log^+ \|\mathbf{C}_0\|)$ and $E(\log^+ \|\mathbf{U}_0\|)$ are finite and therefore, the process (8) has a strictly stationary solution that is given by

$$\mathbf{Z}_t = \sum_{k=0}^{\infty} E\left(\mathbf{C}_t \mathbf{C}_{t-1} \ldots \mathbf{C}_{t-k+1}\right) \diamond \mathbf{U}_{t-k}. \tag{26}$$

Conversely, let assume that there exists a strictly stationary solution $\{\mathbf{Z}_t\}_{t\in Z}$ of Eq. (6). By iterating Eq. (8), we have for $t > 0$,

$$\begin{aligned}
\mathbf{Z}_0 &= E\left(\mathbf{C}_0\right) \diamond \mathbf{Z}_{-1} + \mathbf{U}_0 \\
&= E\left(\mathbf{C}_0 \mathbf{C}_{-1}\right) \diamond \mathbf{Z}_{-2} + \mathbf{U}_0 + E\left(\mathbf{C}_0\right) \diamond \mathbf{U}_{-1} \\
&= E\left(\mathbf{C}_0 \mathbf{C}_{-1} \ldots \mathbf{C}_{-t}\right) \diamond \mathbf{Z}_{-t-1} + \mathbf{U}_0 + \sum_{j=1}^{t} E\left(\mathbf{C}_0 \ldots \mathbf{C}_{-j+1}\right) \diamond \mathbf{U}_{-j} \\
&= E\left(\prod_{j=0}^{t} \mathbf{C}_{-j}\right) \diamond \mathbf{Z}_{-t-1} + \sum_{j=0}^{t} E\left(\prod_{i=0}^{j-1} \mathbf{C}_{-i}\right) \diamond \mathbf{U}_{-j} \\
\mathbf{Z}_0 &= E\left(\mathbf{C}^{(t)}\right) \diamond \mathbf{Z}_{-t-1} + \mathbf{U}^{(t)}
\end{aligned}$$

with $t \in N_0$ and where $\prod_{i=0}^{-1} \mathbf{C}_{-i} = 1$.

All the coefficients of $\mathbf{C}_t$, $\mathbf{Y}_t$ and $\mathbf{U}_t$ are nonnegative. The characteristic polynomial of $E(\mathbf{C}_t)$ is $\Psi(z) = z^n - \psi_1 z^{n-1} - \cdots - \psi_{n-1} z - \psi_n$. By Lemma 1, the roots of $\Psi(z)$ are all inside the unit circle, then $\lim_{t\to\infty} E\left(\left(\prod_{j=0}^{t} \mathbf{C}_{-j}\right) e_i\right) = 0$ a.s. where e_i denotes the canonical basis of R^n. We have

$$\lim_{t\to\infty} E\left(\prod_{j=0}^{t} \mathbf{C}_{-j}\right) = 0$$

$$\lim_{t\to\infty} E\left(\| \prod_{j=0}^{t} \mathbf{C}_{-j}\|\right) = 0$$

then $\lim_{t\to\infty} E\left(\prod_{j=0}^{t} \mathbf{C}_{-j}\right) \diamond \mathbf{Y}_{-t-1} = 0$ a.s. According to Lemma 2 the associated top Lyapunov exponent γ is strictly negative, so that the series $\sum_{j=0}^{t} E\left(\prod_{i=0}^{j-1} \mathbf{C}_{-i}\right) \diamond \mathbf{U}_{-j}$ converges a.s.

Therefore, $\{\mathbf{Z}_t\}_{t\in Z}$ is a strictly stationary solution of Eq. (6). Furthermore, we can write the solution in (13) as $\mathbf{Z}_t = F_t(\mathbf{C}_{t-1}, \mathbf{C}_{t-1}, \ldots, \mathbf{U}_t)$ for some measurable function F independent of t. It follows that the strictly stationary solution is also ergodic because $\mathbf{C}_t$ and $\mathbf{U}_t$ are ergodic, see Brandt et al. (1990) Lemma A 1.2.7.

Now, we aim to prove the uniqueness of the strictly stationary solution. Let $\{\mathbf{W}_t\}_{t\in Z}$ be another strictly stationarity solution of Eq. (8). The norm of the following difference for $t > 0$

$$\begin{aligned} \|\mathbf{Z}_0 - \mathbf{W}_0\| &= \|E\left(\mathbf{C}_0\mathbf{C}_{-1}\ldots\mathbf{C}_{-t}\right)\diamond\left(\mathbf{Z}_{-t-1} - \mathbf{W}_{-t-1}\right)\| \\ &\le \|E\left(\mathbf{C}_0\mathbf{C}_{-1}\ldots\mathbf{C}_{-t}\right)\diamond\|\,\|\mathbf{Z}_{-t-1} - \mathbf{W}_{-t-1}\| \\ &\le \|E\left(\mathbf{C}^{(t)}\right)\diamond\|\,\|\mathbf{Z}_{-t-1} - \mathbf{W}_{-t-1}\|, \end{aligned}$$

by Lemma 1, converges to 0, a.s. and the fact that the law of the difference $(\mathbf{Z}_{-t-1} - \mathbf{W}_{-t-1})$ is independent of t, imply that $\mathbf{Z}_0 - \mathbf{W}_0$ converges to 0 in probability. We conclude that $\mathbf{Z}_0 = \mathbf{W}_0$ and that Eq. (6) has a unique solution, once the counting process are known. ■

Lemma 3. *(Lemma 1 in Bakic and Guljas (1999)) Let $A \in B(H)$ be an operator represented as a block matrix*

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \ldots & A_{1,n} & \ldots \\ A_{2,1} & A_{2,2} & \ldots & A_{2,n} & \ldots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_{n,1} & A_{n,2} & \ldots & A_{n,n} & \ldots \\ \vdots & \vdots & & \vdots & \ddots \end{pmatrix}, \tag{27}$$

corresponding to an orthogonal decomposition $H = \bigoplus_{i=1}^{\infty} H_i$. Let us denote

$$R_i = \sum_{j=1}^{\infty} \|A_{i,j}\|, \qquad R = \sup\{R_i : i \in N\} \tag{28}$$

$$C_i = \sum_{i=1}^{\infty} \|A_{i,j}\|, \qquad C = \sup\{C_i : i \in N\}. \tag{29}$$

Then $\|\mathbf{A}\| \le \sqrt{R \cdot C}$.

Proof of Proposition 4. According to Lemma 2 in Bakic and Guljas (1999) the operator $\mathbf{C}$ allows a suitable Hilbert-Schmidt perturbation and thus, has a tri-block matrix representation. With the tri-block matrix representation of $\mathbf{C}$ and Lemma 3 we can provide proofs to Proposition 4 following the proofs of Theorem 2 in Bakic and Guljas (1999).

We first prove the necessity. Let

$$\mathbf{D}_n = \mathbf{C}^*_{n-1,n}\mathbf{C}_{n-1,n} + \mathbf{C}^*_{n,n}\mathbf{C}_{n,n} + \mathbf{C}^*_{n+1,n}\mathbf{C}_{n+1,n} \in L(H_n), \forall n \in N.$$

Then each $\mathbf{D}_n$ is a positive finite rank operator, so there exists a unit vector $x_n \in H_n$ such that $\|\mathbf{D}_n\| = \|\mathbf{D}_n x_n\|$, $\forall\, n$. The orthonormal sequence (x_n) obtained in this way has the property $\|\mathbf{C}_n\|^2 = \|\mathbf{D}_n x_n\| = \|\mathbf{D}_n\|$, $\forall n$, and because $\mathbf{C}$ is compact this implies $\lim_n \|\mathbf{D}_n\| = 0$, hence

$$\lim_n \|\mathbf{C}_{n,n}\| = \lim_n \|\mathbf{C}_{n+1,n}\| = \lim_n \|\mathbf{C}_{n,n+1}\| = 0.$$

Conversely, let us denote by $P_n \in B(H)$ the orthogonal projection of the finite rank onto $\bigoplus_{i=1}^n H_i$. Consequently, $\mathbf{C}_n = P_n \mathbf{C} P_n$ is also a finite rank operator and

$$\mathbf{C} - \mathbf{C}_n = \begin{pmatrix} 0 & 0 & 0 & \dots & & & \\ 0 & \ddots & \ddots & \ddots & & & \\ 0 & \ddots & 0 & 0 & 0 & & \\ \vdots & \ddots & 0 & 0 & \mathbf{C}_{n,n+1} & 0 & \\ & & 0 & \mathbf{C}_{n+1,n} & \mathbf{C}_{n+1,n+1} & \ddots & \ddots \\ & & & 0 & \ddots & \ddots & \\ & & & & \ddots & & \end{pmatrix}. \tag{30}$$

By hypothesis we have $\lim_n \|\mathbf{C}_{n,n}\| = \lim_n \|\mathbf{C}_{n+1,n}\| = \lim_n \|\mathbf{C}_{n,n+1}\| = 0$, so one can find n large enough to make the estimators R and C (see Lemma 3) for the operator $\mathbf{C} - \mathbf{C}_n$ arbitrary small. Then, by Lemma 3, $\|\mathbf{C} - \mathbf{C}_n\| \le \sqrt{R \cdot C}$ and since $\mathbf{C}_n$ has a finite rank, $\mathbf{C}$ must be compact. ■

Proof of Proposition 5. The assumptions (1.1) and (1.2) follow from the square integrability of $\log^+ \|\mathbf{C}\|$ and the ergodic theorem. For $q = 1$, $\lim \frac{1}{t} \log \|(\mathbf{C}^t)\| = l_1$. $\mathbf{C}^t$ acts as a bounded linear operator on $H^{\wedge q}$ and since $\|(\mathbf{C}^t)^{\wedge q}\| \le \|\mathbf{C}^t\|^q$, the assumption (1.3) holds.

Let define

$$\varpi_N = \{x \in \Gamma^+ : \lim_{q \to \infty} \frac{1}{q} l_q^+(x) \ge -N\}.$$

Then

$$\begin{aligned} -N p(\varpi_N) &\le \int_{\varpi_N} p(dx) \frac{1}{q} l_q^+(x) \\ &\le \int_{\varpi_N} p(dx) \frac{1}{q} \log \|\mathbf{C}(x)^{\wedge q}\|. \end{aligned} \tag{31}$$

Because $\mathbf{C}(x)$ is compact, it follows that when $q \to \infty$, then $\frac{1}{q} \log \|\mathbf{C}(x)^{\wedge q}\| \to -\infty$.
Since

$$\frac{1}{q} \log \|\mathbf{C}(\cdot)^{\wedge q}\| \le \log^+ \|\mathbf{C}(\cdot)\| \in L^1(M, p),$$

we must have $p(\varpi_N) = 0$ for all real N. ■

Proof of corollary 2. By induction on n, we have

$$\det(zIn - E(\mathbf{C}_1)) = z^n\left(1 - \sum_{i=1}^{n}\psi_i z^{-i}\right).$$

The inequality $|a-b| \geq |(|a|-|b|)|$ implies that if $|z|>1$, then

$$\det(zI_n - E(\mathbf{C}_1)) > 1 - \sum_{i=1}^{n}\psi_i. \tag{32}$$

Since the right-hand side is zero and since $\det(zI_n - E(\mathbf{C}_1)) = 0$, we conclude that the spectral radius ρ of the matrix $E(\mathbf{C}_1)$ is 1. Furthermore, all the coefficients of the matrix $\mathbf{C}_2\mathbf{C}_1$ are almost surely positive and $\mathbf{C}_1$ has no zero column nor zero row. Since $\mathbf{C}_1$ is not a.s. bounded, these properties imply by theorem 2 in Kesten and Spitzer (1984) that the top Lyapunov exponent γ satisfies $\gamma < \log\rho$. As result, $\gamma < 0$ and the corollary follows from Proposition 5. ■

Proof of Proposition 6. We note that all the elements in $\mathbf{Z}_t$, $\mathbf{C}_t$ and $\mathbf{U}_t$ are strictly positive, thus the model is irreducible, see Bougerol and Picard (1992) for more details.

Now let go back to the model in (8) and show that it has higher order moment: We have

$$\begin{aligned} E\left(\mathbf{Z}_t^{\otimes m}\right) &\geq E\left(E(\mathbf{C}_t)\diamond\mathbf{Z}_{t-1}\right)^{\otimes m} + E\left(\mathbf{U}_t^{\otimes m}\right)\\ &= E\left(\mathbf{C}_t\right)^{\otimes m}E\left(\mathbf{Z}_{t-1}^{\otimes m}\right)G_1R_1^{\otimes m}\\ &\geq G_1\sum_{j=0}^{n}\left[E\left(\mathbf{C}_t\right)^{\otimes m}\right]^j R_1^{\otimes m}\end{aligned} \tag{33}$$

$G_1 = \min\{$all the positive elements of $E\left(\mathbf{U}_i^{\otimes m}\right)\}$, $R_1 = (1,0,0\ldots,0)'$. A vector $A >$ a vector B means that each element of A exceeds the corresponding element of B.

If n tends to infinity, from (33) we have

$$\sum_{j=0}^{n}\left[E\left(\mathbf{C}_t\right)^{\otimes m}\right]^j R_1^{\otimes m} < \infty. \tag{34}$$

The idea here is to make use of the nonnegativity of the elements of $E\left(\mathbf{C}_t\right)^{\otimes m}$ and $R_1^{\otimes m}$. We first show that

$$\left[E\left(\mathbf{C}_t^{\otimes m}\right)\right]^Q R_1^{\otimes m} > 0. \tag{35}$$

We will prove that (35) holds. First $E\left(\mathbf{C}_t\right)^{\otimes m}R_1^{\otimes m} = E\left(\mathbf{C}_tR_1\right)^{\otimes m}$, where $\mathbf{C}_tR_1 = (\xi_{t-1}^{(1)}, 1, 0\ldots,0)$. Let $G_2 = \min\{$all the positive elements of $E\left(\mathbf{C}_tR_1\right)^{\otimes m}\}$ and $R_2 = (1,1,0,0,\ldots,0)'$. It follows that

$$E\left(\mathbf{C}_t\right)^{\otimes m}R_1^{\otimes m} \geq G_2R_2^{\otimes m}. \tag{36}$$

From (36), we have

$$\left[E\left(\mathbf{C}_t\right)^{\otimes m}\right]^2 R_1^{\otimes m} \geq G_2E\left(\mathbf{C}_t\right)^{\otimes m}R_2^{\otimes m} = G_2E\left(\mathbf{C}_tR_2\right)^{\otimes m}. \tag{37}$$

Now, $\mathbf{C}_t R_2 = (\xi_{t-1}^{(1)} + \xi_{t-2}^{(2)}, 1, 1, 0 \ldots, 0)$. Let $G_3 = \min\{\text{all the positive elements of } E(\mathbf{C}_t R_2)^{\otimes m}\}$ and $R_3 = (1, 1, 1, 0, \ldots, 0)'$. From (37), we have

$$\left[E(\mathbf{C}_t)^{\otimes m}\right]^2 R_1^{\otimes m} \geq G_2 G_3 R_3^{\otimes m}. \tag{38}$$

Repeating the preceding procedure Q times, we can show that

$$\left[E(\mathbf{C}_t)^{\otimes m}\right]^Q R_1^{\otimes m} \geq \left(\prod_{j=2}^{Q} G_j\right) R_Q^{\otimes m}, \tag{39}$$

where $G_j > 0$ and $R_Q = (1, 1, 1, \ldots, 1)$. Thus, (35) holds. From (34) and (35), we have

$$\sum_{j=0}^{\infty} \left[E(\mathbf{C}_t)^{\otimes m}\right]^j \left[E(\mathbf{C}_t)^{\otimes m}\right]^Q R_1^{\otimes m} < \infty. \tag{40}$$

Let c_{kl}^j be the $(k, l)th$ element of $\left[E(\mathbf{C}_t)^{\otimes m}\right]^j$. From (40), we know that $\sum_{j=0}^{\infty} c_{kl}^j < \infty$ for all $k, j = 1, \ldots, n^m$, i.e.,

$$\sum_{j=0}^{\infty} \left[E(\mathbf{C}_t)^{\otimes m}\right]^j < \infty, \tag{41}$$

and hence $\rho\left[E\left(\mathbf{C}_t^{\otimes m}\right)\right] < 1$. ■

Proof of Proposition 7. The INHYGARCH(p, d, q) is process is strictly stationary and ergodic under Assumption 3 and according to Proposition 5. The first derivative of the logarithm likelihood function with respect to θ is given by:

$$\frac{\partial l_t(\boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\theta}} = \left(\frac{y_t}{\lambda_t(\boldsymbol{\theta})} - 1\right) \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{42}$$

with $\lambda_t = \beta_0 [1 - \beta(1)]^{-1} + \psi(\mathrm{B}) y_t$.

The second derivative of the logarithm likelihood function with respect to θ is:

$$\frac{\partial^2 l_t(\boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \left(\frac{y_t}{\lambda_t(\boldsymbol{\theta})} - 1\right) \frac{\partial^2 \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{y_t}{\lambda_t^2(\boldsymbol{\theta})} \frac{\partial^2 \lambda_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}. \tag{43}$$

For simplicity, we only consider the case where $p = q = 0$, and $\lambda_t(\boldsymbol{\theta}) = \sum_{i=0}^{\infty} a_i(\boldsymbol{\theta}) y_{t-i}$, with $a_i = O(i^{-1-d})$ for INHYGARCH(p, d, q) or INFIGARCH(p, d, q). We note that the general case $p \neq q \neq 0$ can be similarly verified. In the case of short-memory INGARCH(p,q) model, $a_i = O(\rho^{-i})$ for INGARCH(1,1), with $\rho = 1/\beta_1$. We note that $a_0 = 1$.

As Θ is compact, there exist $\underline{d}$ and $\bar{d}$ such that $0 < \underline{d} \leq d \leq \bar{d} < 1$. Thus,

$$\sup_{\boldsymbol{\theta} \in \Theta} |a_i(\boldsymbol{\theta})| = O(i^{-1-\underline{d}}) \tag{44}$$

It follows that

$$\sup_{\boldsymbol{\theta}\in\Theta}|l_t(\boldsymbol{\theta}))| = \sup_{\boldsymbol{\theta}\in\Theta}\left|y_t \ln \sum_{i=0}^{\infty} a_i(\boldsymbol{\theta})y_{t-i} - \sum_{i=0}^{\infty} a_i(\boldsymbol{\theta})y_{t-i}\right| \tag{45}$$
$$\leq |y_t|\,|\ln y_t - 1| + O(1)\sum_{i=1}^{\infty} i^{-1-\underline{d}}\,|y_t \ln y_{t-i} - y_{t-i}|\,.$$

It is obvious that $\mathrm{E}\sup_{\Theta}|l_t(\boldsymbol{\theta}))| < \infty$. Furthermore, $l_t(\boldsymbol{\theta})$ is continuous on Θ and $l_t^{''}(\boldsymbol{\theta}) < 0$ for all $\boldsymbol{\theta}$ that are interior points of Θ. This implies that l is a concave function and that any stationary point of l is the unique global maximum of l. Thus, Assumption 1(i) holds.

We next show that 1(ii) and 1(iii) also hold.

$$\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\left(\frac{y_t}{\lambda_t(\boldsymbol{\theta})} - 1\right) \tag{46}$$

with $\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \ln(1-L)(1-L)^d y_t$.

Using Eq. (46), we have

$$\mathrm{E}\left(D_t(\boldsymbol{\theta}_0)|\Omega_t\right) = \frac{\partial \lambda_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\mathrm{E}\left(\frac{y_t}{\lambda_t(\boldsymbol{\theta}_0)} - 1\right) = 0, \tag{47}$$

and if $d \in (0,1)$, $0 < \Upsilon = \left(\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\right)^2 < \infty$. This implies that $D_t(\boldsymbol{\theta})$ is a martingale difference and Assumption 1(ii) is satisfied.

$$\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \ln(1-L)(1-(1-L)^d)y_t = \sum_{i=1}^{\infty} b_{1i}(d)y_{t-i}. \tag{48}$$

$$\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} = \frac{\partial^2 \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\left(\frac{y_t}{\lambda_t(\boldsymbol{\theta})} - 1\right) - \left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2 \frac{y_t}{\lambda_t^2(\boldsymbol{\theta})}. \tag{49}$$

$$\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} = -\ln^2(1-L)(1-L)^d y_t = \sum_{i=1}^{\infty} b_{2i}(d)y_{t-i}, \tag{50}$$

with $\sup_{\boldsymbol{\theta}\in\Theta}|b_{ki}(\boldsymbol{\theta})| = O(i^{-1-\underline{d}})$ for $k = 1, 2$.

$$\mathrm{E}\left(-\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right) = \mathrm{E}\left[-\frac{\partial^2 \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\left(\frac{y_t}{\lambda_t(\boldsymbol{\theta})} - 1\right)\right] + \mathrm{E}\left[\left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2 \frac{y_t}{\lambda_t^2(\boldsymbol{\theta})}\right] \tag{51}$$
$$= \mathrm{E}\left[\left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2 \frac{y_t}{\lambda_t^2(\boldsymbol{\theta})}\right] > 0,$$

and

$$\mathrm{E}\sup \| \ln^2(1-\mathrm{B})(1-\mathrm{B})^d y_t \| = \mathrm{E}\sup | \sum_{i=1}^{\infty} b_{2i}(\boldsymbol{\theta}) y_{t-i} | \tag{52}$$

$$< c\mathrm{E}\left(\sum_{i=1}^{\infty} i^{-1-d} |y_{t-i}| \right) = O(t^{-d}).$$

Thus, Assumption 1(iii) holds.

We consider Assumption 2. For simplicity, let $\tilde{y}_0 = (0, 0, \dots)$. This leads to $\tilde{l}_t(\boldsymbol{\theta}) = 0$, $\tilde{D}_t(\boldsymbol{\theta}) = 0$ and $\tilde{P}_t(\boldsymbol{\theta}) = 0$. We recall that $\tilde{l}_t(\boldsymbol{\theta})$, $\tilde{D}_t(\boldsymbol{\theta}_0)$ and $\tilde{P}_t(\boldsymbol{\theta})$ represent $l_t(\boldsymbol{\theta})$, $D_t(\boldsymbol{\theta}_0)$ and $P_t(\boldsymbol{\theta})$ evaluated at the initial value $\tilde{y}_0 = (0, 0, \dots)$, respectively. The proof of the Assumption 1(i) also applies to Assumption 2(i). We now show that Assumption 2(ii) holds:

$$\lim_{l\to\infty} \mathrm{P}\left(\max_{l \le n < \infty} \frac{1}{\sqrt{n}} \| \sum_{t=1}^{n} \left[S_t(\boldsymbol{\theta}) - \tilde{S}_t(\boldsymbol{\theta}) \right] \| > \epsilon \right) = \lim_{l\to\infty} \mathrm{P}\left(\max_{l \le n < \infty} \frac{1}{\sqrt{n}} \| \sum_{t=1}^{n} \ln(1-\mathrm{B})(1-(1-\mathrm{B})^d) y_t \| > \epsilon \right) \tag{53}$$

$$= \lim_{l\to\infty} \mathrm{P}\left(\max_{l \le n < \infty} \frac{1}{\sqrt{n}} \| \left[\sum_{k=1}^{\infty} b_{1i} y_{t-k} \right] \| > \epsilon \right)$$

$$= \lim_{l\to\infty} \mathrm{P}\left(\max_{l \le n < \infty} \frac{1}{\sqrt{n}} \| \sum_{t=1}^{n} \left[\sum_{k=1}^{\infty} b_{1i} y_{t-k} \right] \| > \epsilon \right)$$

$$= 0.$$

From the theory of infinite sums, $\ln(1-z) = -\sum_{k=1}^{\infty} \frac{z^k}{n}$ converges a.s. if $|z| < 1$. Similar to the proof of the Assumption 1(iii), one can also show the proof for Assumption 2(iii).

■

6 Conclusion

We have shown that the unique strictly stationary solution for the INARCH(∞) processes is ergodic, but not geometric ergodic. Geometric ergodicity of INARCH models is crucial for proving asymptotic normality of the conditional maximum likelihood estimates of the parameters in models. We believe that the proofs for geometric ergodicity can obtained based on the ergodicity results in Bhattacharya and Chanho (1995) for nonlinear first order autoregressive models.

References

Bakic, D. and B. Guljas (1999). A note on compact operators and operator matrices. *Mathematical Communications 4*, 159–165.

Bhattacharya, R. N. and L. Chanho (1995). Ergodicity of nonlinear first order autoregressive models. *Journal of Theoretical Probability 8*, 207–219.

Bougerol, P. and N. Picard (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics 52*, 115–127.

Brandt, A., P. Franken, and B. Lisek (1990). *Stationary Stochastic Models*. Wiley, Chichester.

Brockwell, P. and R. Davis (1991). *Time Series: Theory and Methods*. 2nd edn. New York: Springer-Verlag.

Doukhan, P., K. Fokianos, and D. Tjøstheim (2012). On weak dependence conditions for Poisson autoregressions. *Bernoulli 82*, 942–948.

Doukhan, P. and O. Wintenberger (2008). Weakly dependent chains with infinite memory. *Stochastic Processes and Their Applications 118*, 1997–2013.

Ferland, R., A. Latour, and D. Oraichi (2006). Integer-valued GARCH process. *Journal of Time Series Analysis 27*, 923–942.

Fokianos, K., A. Rahbek, and D. Tjøstheim (2009). Poisson autoregression. *Journal of the American Statistical Association 104*, 1430–1439.

Fokianos, K. and D. Tjøstheim (2012). Nonlinear Poisson autoregression. *Annals of the Institute of Statistical Mathematics 64*, 1205–1225.

Kazakevicius, V. and R. Leipus (2002). On stationarity in the ARCH(∞) model. *Econometric Theory 18*, 1–16.

Kedem, B. and Fokianos (2002). *Regression Models for Time Series Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken.

Kesten, H. and F. Spitzer (1984). Convergence in distribution of products of random matrices. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 67*, 363–386.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis 4*, 439–455.

Ling, S. and M. McAleer (2010). A general asymptotic theory for time-series models. *Statistica Neerlandica 64*, 97–111.

Neumann, M. H. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli 17*, 1268–1284.

Oseledec, V. (1968). A multiplicative ergodic theorem. *Trans. Moscow Math. Soc. 19*, 197–231.

Ruelle, D. (1982). Characteristic exponents and invariant manifolds in hilbert space. *Annals of Mathematics 115*, 243–290.

Schaumlöffel, K. (1991). Multiplicative ergodic theorems in infinite dimensions. In L. Arnold, H. Crauel, and J. Eckmann (Eds.), *Lyapunov Exponents*, Volume 1486, Chapter Lecture Notes in Mathematics. Springer, Berlin, Heidelberg. https://doi.org/10.1007/BFb0086668.

Segnon, M. and M. Stapper (2019). Long memory conditional heteroscedasticity in count data. Working Paper 82/2019, University Muenster.

Sim, T., R. Douc, and F. Roueff (2021). General-order observation-driven models: Ergodicity and consistency of the maximum likelihood estimator. *Electronic Journal of Statistics 15*, 3349–3393.

Van Barel, M., V. Ptak, and Z. Vavrin (1999). Extending the notions of companion and infinite companion to matrix polynomials. *Linear Algebra and its Applications 290*, 61–94.

Weiß, C. H. (2008). Thinning operations for modeling time series of counts - survey. *AStA Advances in Statistical Analysis 92*, 319–341.

Zerner, M. P. W. (2018). Recurrence and transience of contractive autoregressive processes and related Markov chains. *Electronic Journal of Probability 23*, 1–24.

Table 2: Simulation Results - Mean point estimates and standard deviations (in parentheses)

Model	d	n	β_0	$\hat{\beta}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{d}$	$\hat{\eta}$
INFIGARCH	0.2	10000	1.00	1.0718	0.2080	0.4954	0.1915	-
				(0.2208)	(0.0314)	(0.0333)	(0.0308)	
	0.2	20000	1.00	1.0331	0.2034	0.4979	0.1962	-
				(0.1521)	(0.0226)	(0.0230)	(0.0227)	
	0.2	50000	1.00	1.0051	0.2003	0.5000	0.1996	-
				(0.0940)	(0.0142)	(0.0148)	(0.0141)	
	0.4	10000	0.25	0.2660	0.2048	0.4958	0.3945	-
				(0.0568)	(0.0310)	(0.0311)	(0.0313)	
	0.4	20000	0.25	0.2556	0.2018	0.4997	0.3975	-
				(0.0367)	(0.0213)	(0.0224)	(0.0213)	
	0.4	50000	0.25	0.2515	0.2003	0.5001	0.3994	-
				(0.0224)	(0.0131)	(0.0138)	(0.0131)	
	0.6	10000	0.04	0.0459	0.2009	0.4974	0.5998	-
				(0.0143)	(0.0296)	(0.0316)	(0.0309)	
	0.6	20000	0.04	0.0427	0.2002	0.4995	0.6001	-
				(0.0080)	(0.0221)	(0.0223)	(0.0227)	
	0.6	50000	0.04	0.0409	0.2000	0.5000	0.6001	-
				(0.0041)	(0.0133)	(0.0142)	(0.0135)	
INHYGARCH	0.2	10000	1.75	2.4720	0.1343	0.4335	0.3574	0.7464
				(0.9992)	(0.1629)	(0.1741)	(0.2562)	(0.2042)
	0.2	20000	1.75	2.3022	0.1491	0.4453	0.3127	0.7750
				(0.9132)	(0.1336)	(0.1578)	(0.2129)	(0.1664)
	0.2	50000	1.75	2.0804	0.1642	0.4730	0.2730	0.7907
				(0.6984)	(0.1037)	(0.1158)	(0.1551)	(0.1216)
	0.4	10000	1.00	1.0440	0.2015	0.5143	0.4174	0.8266
				(0.2107)	(0.0660)	(0.0795)	(0.1030)	(0.0947)
	0.4	20000	1.00	1.0209	0.2021	0.5064	0.4041	0.8431
				(0.1525)	(0.0516)	(0.0585)	(0.0738)	(0.0598)
	0.4	50000	1.00	1.0123	0.2025	0.5037	0.4012	0.8451
				(0.1053)	(0.0388)	(0.0480)	(0.0572)	(0.0440)
	0.6	10000	0.80	0.8066	0.2251	0.4854	0.5732	0.8442
				(0.1137)	(0.0860)	(0.838)	(0.1006)	(0.0330)
	0.6	20000	0.80	0.8071	0.2203	0.4867	0.5783	0.8463
				(0.0753)	(0.0667)	(0.0629)	(0.0774)	(0.0220)
	0.6	50000	0.80	0.8086	0.2187	0.4849	0.5790	0.8477
				(0.0503)	(0.0494)	(0.0437)	(0.0565)	(0.0151)