



HAL
open science

GENERAL-ORDER OBSERVATION-DRIVEN MODELS

Tepmony Sim, Randal Douc, François Roueff

► **To cite this version:**

Tepmony Sim, Randal Douc, François Roueff. GENERAL-ORDER OBSERVATION-DRIVEN MODELS. 2016. hal-01383554v1

HAL Id: hal-01383554

<https://hal.science/hal-01383554v1>

Preprint submitted on 18 Oct 2016 (v1), last revised 7 Jun 2021 (v3)

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

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



Distributed under a Creative Commons Attribution 4.0 International License

GENERAL-ORDER OBSERVATION-DRIVEN MODELS

BY TEPMONY SIM^{*}, RANDAL DOUC[†] AND FRANÇOIS ROUEFF[‡]

Institute of Technology of Cambodia^{}, Télécom SudParis[†] and Télécom ParisTech[‡]*

The class of observation-driven models (ODMs) includes the GARCH(1, 1) model as well as integer-valued time series models such as the log-linear Poisson GARCH of order (1, 1) and the NBIN-GARCH(1, 1) models. In this contribution, we treat the case of general-order ODMs in a similar fashion as the extension of the GARCH(1, 1) model to the GARCH(p, q) model. More precisely, we establish the stationarity and the ergodicity as well as the consistency and the asymptotic normality of the maximum likelihood estimator (MLE) for the class of general-order ODMs, under conditions which are easy to verify. We illustrate these results with specific observation-driven time series, namely, the log-linear Poisson GARCH of order (p, q) and the NBIN-GARCH(p, q) models. An empirical study is also provided.

1. Introduction. Since the introduction by [Cox \(1981\)](#), observation-driven models (ODMs) have been receiving renewed interest in recent years. These models are widely applied in various fields ranging from economics (see [Pindyck and Rubinfeld \(1998\)](#)), environmental study (see [Bhaskaran et al. \(2013\)](#)), epidemiology and public health study (see [Zeger \(1988\)](#); [Davis, Dunsmuir and Wang \(1999\)](#); [Ferland, Latour and Oraichi \(2006\)](#)), finance (see [Liesenfeld and Richard \(2003\)](#); [Rydberg and Shephard \(2003\)](#); [Fokianos and Tjøstheim \(2011\)](#); [Francq and Zakoian \(2011\)](#)) and population dynamics (see [Ives et al. \(2003\)](#)). The celebrated GARCH(1, 1) model introduced in [Bollerslev \(1986\)](#) as well as most of the models derived from this one are typical examples of ODMs; see [Bollerslev \(2008\)](#) for a list of some of them. A list of contributions on this class of models specifically dealing with discrete data includes [Streett \(2000\)](#); [Davis, Dunsmuir and Streett \(2003\)](#); [Heinen \(2003\)](#); [Ferland, Latour and Oraichi \(2006\)](#); [Fokianos, Rahbek and Tjøstheim \(2009\)](#); [Franke \(2010\)](#); [Fokianos and Tjøstheim \(2011\)](#); [Henderson, Matteson and Woodard \(2011\)](#); [Neumann \(2011\)](#); [Davis and Liu \(2012\)](#); [Doukhan, Fokianos and Tjøstheim \(2012\)](#); [Douc, Doukhan and Moulines \(2013\)](#); [Fokianos et al. \(2013\)](#); [Christou and Fokianos \(2014, 2015\)](#)

MSC 2010 subject classifications: Primary 60J05, 62F12; secondary 62M05, 62M10

Keywords and phrases: consistency, ergodicity, general-order, maximum likelihood, observation-driven models, time series of counts

and [Douc, Roueff and Sim \(2015\)](#).

ODMs have the nice feature that the computations of the associated (conditional) likelihood and its derivatives are easy, the parameter estimation is hence relatively simple, and the prediction, which is a prime objective in many time series applications, is straightforward. However, it turns out that the asymptotic properties of the maximum likelihood estimator (MLE) for this class can be cumbersome to establish, except when they can be derived using computations specific to the studied model (the GARCH(1, 1) case being one of the most celebrated example). The literature concerning the asymptotic theory of the MLE when the observed variable has Poisson distribution includes [Fokianos, Rahbek and Tjøstheim \(2009\)](#); [Fokianos and Tjøstheim \(2011, 2012\)](#) and [Wang et al. \(2014\)](#). For a more general case where the model belongs to the class of one-parameter exponential ODMs, such as the Bernoulli, the exponential, the negative binomial (with known frequency parameter) and the Poisson autoregressive models, the consistency and the asymptotic normality of the MLE have been derived in [Davis and Liu \(2012\)](#). However, the one-parameter exponential family is inadequate to deal with models such as multi-parametric, mixture or multivariate ODMs (the negative binomial with all unknown parameters and mixture Poisson ODMs are examples of this case). A more general consistency result, yet not the asymptotic normality, has been obtained recently in [Douc, Doukhan and Moulines \(2013\)](#). This general result allows the observed process to admit various forms of distribution and to take values in any Borel space, and allows the hidden process to assume values in any locally compact Polish space endowed with the associated Borel σ -field. This result has later been extended and refined in [Douc, Roueff and Sim \(2015\)](#). However, most of the results obtained so far have been derived only under the framework of GARCH(1, 1)-type or first-order ODMs, yet less is known for the GARCH(p, q)-type discrete ODMs, as highlighted as a remaining unsolved problem in [Tjøstheim \(2015\)](#).

In this contribution, we consider among others (see [Streett \(2000\)](#); [Heinen \(2003\)](#)) a general class of ODMs that is capable to account for several lagged variables of both hidden and observation processes. Namely, we develop theory and inference for the class of general-order ODMs parallel to the GARCH(p, q) family, under the assumption of well-specified models. For the development on the GARCH(p, q) model, see for example [Francq and Zakoian \(2004, 2011\)](#) and for multivariate case, see [Comte and Lieberman \(2003\)](#). We extend the approaches introduced in [Douc, Doukhan and Moulines \(2013\)](#) and later used in [Douc, Roueff and Sim \(2015\)](#) to establish the consistency of the MLE and the ergodicity for the models in this

general context. These results can in principle be obtained by embedding the studied model into the corresponding first-order one and then applying the results obtained therein to the embedded model. By appropriate adjustments, the consistency of the MLE for the general-order ODM can be derived. Yet for ergodicity, the generalized result seems not trivial; it turns out that more general assumptions are required, compared to the usual first-order ODMs. In either [Douc, Doukhan and Moulines \(2013\)](#) or [Douc, Roueff and Sim \(2015\)](#), the ergodicity of the usual first-order ODMs is obtained by showing the uniqueness of invariant probability measure for the hidden process, which is by its own right a Markov chain. In these papers, showing the existence of the invariant probability measure for the hidden process relies on the Foster-Lyapunov-type assumption; however, for general ODMs or embedded ones, this assumption may be replaced by the *iterative* Foster-Lyapunov-type assumption instead (see Assumption (AG-4) of [Section 3.4](#)). The asymptotic normality of the MLE is also investigated in this general setting. In this study, we restrict our consideration to the special yet important case where the current hidden state variable is expressed as a linear function with respect to its own past variables. This class is rich enough to cover most of the familiar ODMs, such as the GARCH(p, q), the log-linear Poisson autoregression of order (p, q) and the NBIN-GARCH(p, q) models, and other instances such as multi-parametric, mixture or multivariate ODMs. To establish the asymptotic normality, we follow the classical approach by first approximating the score function by the stationary version of it and then developing the stationary score function around the true parameter through a Taylor expansion. By appropriate assumptions, the central limit theorem for Martingale difference applies and the stationary score function can then be shown to be asymptotically Gaussian. Then assuming invertibility of the asymptotic Fisher information matrix, the asymptotic normality follows. All the results are presented under sufficient and easy-to-check conditions. As a demonstration, they are then applied to the log-linear Poisson autoregression of order (p, q) and the NBIN-GARCH(p, q) models. Finally, we provide an empirical study suggesting that in some circumstance higher-order model may fit the data better than the first-order one if the class of ODMs is used.

This paper is structured as follows. Specific definitions and notation are introduced in [Section 2](#). [Section 3](#) presents the main results of consistency and asymptotic normality of the maximum likelihood estimator (MLE) as well as ergodic solution for the model. We apply these results to some specific models in [Section 4](#). Numerical experiments is given in [Section 5](#). Finally, [Section 6](#) contains the postponed proofs.

2. Definitions and Notation. Let $(\mathsf{X}, d_{\mathsf{X}})$ be a complete and separable metric space endowed with the associated Borel σ -field \mathcal{X} and $(\mathsf{Y}, \mathcal{Y})$ be a Borel space. Let (Θ, Δ) , the set of parameters, be a compact metric space equipped with the metric Δ , $\{G^\theta : \theta \in \Theta\}$ be a family of probability kernels on $\mathsf{X} \times \mathsf{Y}$ and p, q be two positive integers. The observation-driven time series model can be formally defined as follows.

DEFINITION 1. A time series $\{Y_k : k \geq -q+1\}$ valued in Y is said to be distributed according to an *observation-driven model of order (p, q)* (hereafter, ODM(p, q)) with parameter $\theta \in \Theta$ if there exist a family of measurable functions $\{(x_{1:p}, y_{1:q}) \mapsto \psi_{y_{1:q}}^\theta(x_{1:p}) : \theta \in \Theta\}$ from $(\mathsf{X}^p \times \mathsf{Y}^q, \mathcal{X}^{\otimes p} \otimes \mathcal{Y}^{\otimes q})$ to $(\mathsf{X}, \mathcal{X})$ and a process $\{X_k : k \geq -p+1\}$ on $(\mathsf{X}, \mathcal{X})$ such that for all $k \in \mathbb{Z}_+$,

$$(2.1) \quad \begin{aligned} X_{k+1} &= \psi_{Y_{k-q+1:k}}^\theta(X_{k-p+1:k}), \\ Y_{k+1} | \mathcal{F}_k &\sim G^\theta(X_{k+1}; \cdot), \end{aligned}$$

where $\mathcal{F}_k = \sigma(X_{-p+1:k+1}, Y_{-q+1:k})$ and $u_{\ell:m} := (u_\ell, \dots, u_m)$ for $\ell \leq m$. Moreover, we say that the model is dominated by some σ -finite measure ν on $(\mathsf{Y}, \mathcal{Y})$ if for all $x \in \mathsf{X}$, the probability kernel $G^\theta(x; \cdot)$ is dominated by ν . In this case we denote by $g^\theta(x; \cdot)$ its Radon–Nikodym derivative, $g^\theta(x; y) = \frac{dG^\theta(x; \cdot)}{d\nu}(y)$, and we always assume that for all $(x, y) \in \mathsf{X} \times \mathsf{Y}$ and for all $\theta \in \Theta$,

$$g^\theta(x; y) > 0.$$

One of the most popular examples of this class is the general GARCH(p, q) model introduced by [Bollerslev \(1986\)](#), where $\mathsf{X} = (0, \infty)$, $\mathsf{Y} = \mathbb{R}$, $G^\theta(x; \cdot)$ is the centered Gaussian distribution of variance x , the deterministic function $\psi_{y_{1:q}}^\theta(x_{1:p}) = \omega + \sum_{i=1}^p a_i x_i + \sum_{i=1}^q b_i y_i^2$ and $\theta = (\omega, a_{1:p}, b_{1:q})$ with $\omega > 0$ and $a_{1:p}, b_{1:q} \geq 0$. This model was then extensively studied by, for example, [Bougerol and Picard \(1992\)](#); [Francq and Zakoian \(2004\)](#); [Francq and Zakoian \(2009\)](#); [Lindner \(2009\)](#); [Francq and Zakoian \(2011\)](#) and the references therein. For other GARCH examples of this class, see [Bollerslev \(2008\)](#).

REMARK 1. When $p = q = 1$, then the ODM(p, q) defined by (2.1) collapses to the first-order ODM considered in [Douc, Doukhan and Moulines \(2013\)](#) and [Douc, Roueff and Sim \(2015\)](#). Note also that setting $r := \max(p, q)$ and defining $\varphi^\theta : \mathsf{X}^r \times \mathsf{Y}^r \rightarrow \mathsf{X}$ by, for all $(x_{-r+1:0}, y_{-r+1:0}) \in \mathsf{X}^r \times \mathsf{Y}^r$,

$$\varphi_{y_{-r+1:0}}^\theta(x_{-r+1:0}) := \psi_{y_{-q+1:0}}^\theta(x_{-p+1:0}),$$

then the $\text{ODM}(p, q)$ can be generally embedded in an $\text{ODM}(r, r)$. Thus without loss of generality, we can always assume that $p = q$.

The inference about the model parameter is performed by relying on the conditional likelihood of the observations (Y_1, \dots, Y_n) given $(X_{-p+1:0}, Y_{-q+1:0}) = (x_{-p+1:0}, y_{-q+1:0}) := z$ for an arbitrary $z \in \mathbf{X}^p \times \mathbf{Y}^q$. The corresponding conditional density function with respect to $\nu^{\otimes n}$ is, under parameter $\theta \in \Theta$, for all $z = (x_{-p+1:0}, y_{-q+1:0}) \in \mathbf{X}^p \times \mathbf{Y}^q$,

$$(2.2) \quad y_{1:n} \mapsto \prod_{k=0}^{n-1} g^\theta \left(\psi^\theta \langle y_{-q+1:k} \rangle (x_{-p+1:0}); y_{k+1} \right),$$

where, for any vector $y_{0:k}$, the function $\psi^\theta \langle y_{0:k} \rangle : \mathbf{X}^p \times \mathbf{Y}^q \rightarrow \mathbf{X}$ is successively defined by, for all $k \in \mathbb{Z}_+$,

$$(2.3) \quad \begin{aligned} & \psi^\theta \langle y_{-q+1:k} \rangle (x_{-p+1:0}) \\ & := \psi_{y_{k-q+1:k}}^\theta \left(\psi^\theta \langle y_{-q+1:k-p} \rangle (x_{-p+1:0}), \dots, \psi^\theta \langle y_{-q+1:k-1} \rangle (x_{-p+1:0}) \right) \end{aligned}$$

with the convention, for $1 \leq j \leq p$,

$$\psi^\theta \langle y_{-q+1:-j} \rangle (x_{-p+1:0}) := x_{-j+1}.$$

Then, the corresponding (conditional) maximum likelihood estimator (MLE) $\hat{\theta}_{z,n}$ of the parameter θ , with $z = (x_{-p+1:0}, y_{-q+1:0})$, is defined by

$$(2.4) \quad \hat{\theta}_{z,n} \in \operatorname{argmax}_{\theta \in \Theta} \mathbf{L}_{z,n}^\theta,$$

where

$$(2.5) \quad \mathbf{L}_{z,n}^\theta := n^{-1} \sum_{k=0}^{n-1} \ln g^\theta \left(\psi^\theta \langle y_{-q+1:0}, Y_{1:k} \rangle (x_{-p+1:0}); Y_{k+1} \right).$$

In this contribution, we investigate the convergence of $\hat{\theta}_{z,n}$ as $n \rightarrow \infty$ for some (well-chosen) value of z under the assumption that the model is well specified and the observations are in a steady state. That is, we assume that the observations $\{Y_k : k \in \mathbb{Z}_+\}$ are distributed according to $\tilde{\mathbb{P}}^{\theta_\star}$ with some $\theta_\star \in \Theta$, where, for all $\theta \in \Theta$, $\tilde{\mathbb{P}}^\theta$ denotes the stationary distribution of the observation-driven time series corresponding to the parameter θ . However, whether such a distribution is well defined is not always obvious. Let us now detail how this probability distribution can be obtained.

For simplicity, we assume that $p = q$ and that $p \geq 2$ and let $Z = X^p \times Y^{p-1}$ and $\mathcal{Z} = \mathcal{X}^{\otimes p} \otimes \mathcal{Y}^{\otimes(p-1)}$. Let $\Psi^\theta : Z \times Y \rightarrow Z$ be defined by, for all $z = (z_{1:2p-1}) \in Z$ and $y \in Y$,

$$(2.6) \quad \Psi_y^\theta(z) := \left(z_2, \dots, z_p, \psi_{z_{p+1:2p-1}, y}^\theta(z_{1:p}), z_{p+2}, \dots, z_{2p-1}, y \right)$$

The function Ψ^θ is indeed measurable on $(Z \times Y, \mathcal{Z} \otimes \mathcal{Y})$. Moreover, for any vector $y_{1:k} \in Y^k$, we define $\Psi^\theta \langle y_{1:k} \rangle$ as a $Z \rightarrow Z$ function obtained by the successive composition of $\Psi_{y_1}^\theta, \Psi_{y_2}^\theta, \dots$, and $\Psi_{y_k}^\theta$,

$$(2.7) \quad \Psi^\theta \langle y_{1:k} \rangle = \Psi_{y_k}^\theta \circ \Psi_{y_{k-1}}^\theta \circ \dots \circ \Psi_{y_1}^\theta .$$

with the convention $\Psi^\theta \langle y_{1:0} \rangle(z) = z$. By letting $Z_k = (X_{k-p+1:k}, Y_{k-p+1:k-1})$ and observing that $\mathcal{F}_k = \sigma(Z_{0:k+1}, Y_{0:k})$, Model (2.1) can be replaced by: for all $k \in \mathbb{Z}_+$,

$$(2.8) \quad \begin{aligned} Z_{k+1} &= \Psi_{Y_k}^\theta(Z_k), \\ Y_{k+1} \mid \mathcal{F}_k^{Z,Y} &\sim H^\theta(Z_{k+1}; \cdot), \end{aligned}$$

where, for all $z = (z_{1:2p-1}) \in Z$,

$$(2.9) \quad H^\theta(z; \cdot) := G^\theta(\Pi_p(z); \cdot)$$

and, for all $j \in \{1, \dots, 2p-1\}$, $\Pi_j(z) = z_j$. By this representation, the ODM(p, p) is thus imbedded in an ODM(1, 1). This in principle allows us to apply the same results obtained for the class of ODMs(1, 1) to the broader class of ODMs(p, p). As an ODM(1, 1), the bivariate process $\{(Z_k, Y_k) : k \in \mathbb{Z}_+\}$ is a Markov chain on the space $(Z \times Y, \mathcal{Z} \otimes \mathcal{Y})$ with transition kernel K^θ satisfying, for all $(z, y) \in Z \times Y$, $A \in \mathcal{Z}$ and $B \in \mathcal{Y}$,

$$(2.10) \quad K^\theta((z, y); A \times B) = \iint \mathbb{1}_{A \times B}(z', y') \delta_{\Psi_y^\theta(z)}(dz') G^\theta(\Pi_p(z); dy').$$

Note also that, by itself, the process $\{Z_k : k \in \mathbb{Z}_+\}$ is a Markov chain on (Z, \mathcal{Z}) with transition kernel R^θ satisfying, for all $z \in Z$ and $A \in \mathcal{Z}$,

$$(2.11) \quad R^\theta(z; A) = \int \mathbb{1}_A(\Psi_y^\theta(z)) H^\theta(z; dy) = \int \mathbb{1}_A(\Psi_y^\theta(z)) G^\theta(\Pi_p(z); dy).$$

Nevertheless, to apply known results to this embedded ODM(1, 1), some generalizations are needed. First we assume that (Y, d_Y) is a locally compact,

complete and separable space equipped with the metric d_Y . Then, since (X, d_X) is a locally compact, complete and separable space, this implies that the space (Z, d_Z) is also locally compact, complete and separable with the metric d_Z appropriately defined as a function of d_X and d_Y so that \mathcal{Z} is the associated Borel σ -field. Throughout this paper, let us assume the following ergodicity assumption.

(AG-1) For all $\theta \in \Theta$, the transition kernel K^θ of the complete chain admits a unique stationary distribution π^θ on $Z \times Y$.

With this assumption, we can now define $\tilde{\mathbb{P}}^\theta$. The following notation and definitions will be used throughout the paper.

DEFINITION 2. Under Assumption (AG-1), we denote by π_1^θ and π_2^θ the marginal distributions of π^θ on Z and Y , respectively, and by \mathbb{P}^θ and $\tilde{\mathbb{P}}^\theta$ the probability distributions defined respectively as follows.

- a) \mathbb{P}^θ denotes the extension of $\mathbb{P}_{\pi^\theta}^\theta$ on the whole line $(Z \times Y)^{\mathbb{Z}}$.
- b) $\tilde{\mathbb{P}}^\theta$ is the corresponding projection on the component $Y^{\mathbb{Z}}$.

We also use the symbols \mathbb{E}^θ and $\tilde{\mathbb{E}}^\theta$ to denote the expectations corresponding to \mathbb{P}^θ and $\tilde{\mathbb{P}}^\theta$, respectively. Moreover, for all $\theta, \theta' \in \Theta$, we write $\theta \sim \theta'$ if and only if $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta'}$. This defines an equivalence relation on the parameter set Θ and the corresponding equivalence class of θ is denoted by $[\theta] := \{\theta' \in \Theta : \theta \sim \theta'\}$.

The equivalence relationship \sim was introduced by [Leroux \(1992\)](#) as an alternative to the classical identifiability condition.

For any probability distribution ξ on $Z \times Y$, we denote by \mathbb{P}_ξ^θ the distribution of the Markov chain $\{(Z_k, Y_k) : k \in \mathbb{Z}_+\}$ with kernel K^θ and initial probability measure ξ .

3. Main Results.

3.1. *Preliminaries.* In well-specified setting, a general result on the consistency of the MLE for a class of first-order ODMs has been obtained in [Douc, Doukhan and Moulines \(2013\)](#). This result has been later extended in [Douc, Roueff and Sim \(2015\)](#) to a wider class of first-order ODMs, but the latter has been derived under a handy set of assumptions which appear to be quite direct and easy when checking. The approach used to establish the convergence of the MLE $\hat{\theta}_{z,n}$ in these references and in our contribution is briefly described as follows.

First, we establish that, as the number of observations $n \rightarrow \infty$, the normalized log-likelihood $L_{z,n}^\theta$ defined in (2.5), for some well-chosen $z \in Z$, can be approximated by

$$n^{-1} \sum_{k=1}^n \ln p^\theta(Y_k | Y_{-\infty:k-1}),$$

where $p^\theta(\cdot|\cdot)$ is a $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. finite real-valued measurable function defined on $(Y^{\mathbb{Z}}, \mathcal{Y}^{\otimes \mathbb{Z}})$. To define $p^\theta(\cdot|\cdot)$, we usually set, for all $y_{-\infty:1} \in Y^{\mathbb{Z}^-}$, whenever the following limit is well defined,

$$(3.1) \quad p^\theta(y_1 | y_{-\infty:0}) = \begin{cases} \lim_{m \rightarrow \infty} g^\theta \left(\Pi_p(\Psi^\theta \langle y_{-m:0} \rangle(z)); y_1 \right) & \text{if the limit exists,} \\ \infty & \text{otherwise.} \end{cases}$$

By (AG-1), the observed process $\{Y_k : k \in \mathbb{Z}\}$ is ergodic under $\tilde{\mathbb{P}}^{\theta_\star}$ and provided that

$$\tilde{\mathbb{E}}^{\theta_\star} \left[\ln^+ p^\theta(Y_1 | Y_{-\infty:0}) \right] < \infty,$$

it then follows that

$$\lim_{n \rightarrow \infty} L_{z,n}^\theta = \tilde{\mathbb{E}}^{\theta_\star} \left[\ln p^\theta(Y_1 | Y_{-\infty:0}) \right], \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Finally, we show that with probability tending to one, the MLE $\hat{\theta}_{z,n}$ eventually lies in a neighborhood of the set

$$(3.2) \quad \Theta_\star = \operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} \left[\ln p^\theta(Y_1 | Y_{-\infty:0}) \right],$$

which only depends on θ_\star , establishing that

$$(3.3) \quad \lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{z,n}, \Theta_\star) = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

where Δ is the metric endowing the parameter space Θ .

In Douc, Roueff and Sim (2015), their easy-to-check conditions also induce that for all $\theta, \theta_\star \in \Theta$, there exists a $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. finite measurable function $\Psi^\theta \langle \cdot \rangle : Z^{\mathbb{Z}^-} \rightarrow Z$ such that for some appropriate value $z \in Z$,

$$(3.4) \quad \lim_{m \rightarrow \infty} \Psi^\theta \langle Y_{-m:0} \rangle(z) = \Psi^\theta \langle Y_{-\infty:0} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

$$(3.5) \quad Z_1 = \Psi^{\theta_\star} \langle Y_{-\infty:0} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

and that

$$(3.6) \quad p^\theta(Y_1 | Y_{-\infty:0}) = g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1 \right), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

In addition, it is shown that for all $\theta, \theta_\star \in \Theta$,

- (i) If $\theta \neq \theta_*$, $y \mapsto p^\theta(y|Y_{-\infty:0})$ is a density function $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.
- (ii) Under $\tilde{\mathbb{P}}^{\theta_*}$, the function $y \mapsto p^{\theta_*}(y|Y_{-\infty:0})$ is the conditional density function of Y_1 given $Y_{-\infty:0}$.

Under these same conditions, [Douc, Roueff and Sim \(2016\)](#) further showed that the maximizing set Θ_* defined in (3.2) indeed reduces to the equivalence class of θ_* through the equivalence relation defined in [Definition 2](#). Then, the convergence in (3.3) yields a so-called *equivalence-class consistency*:

$$(3.7) \quad \lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{z_1, n}, [\theta_*]) = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.},$$

for some well-chosen $z_1 \in \mathbf{Z}$. From (3.7) the (strong) consistency of the MLE $\hat{\theta}_{z, n}$ will immediately follow if one can show that $[\theta_*]$ reduces to the singleton $\{\theta_*\}$, which is often referred to as solving *identifiability* problem. In this contribution, this issue is also investigated. We will show that if moreover the probability density kernel G^θ and the measurable function $\Psi^\theta \langle Y_{-\infty:0} \rangle$ defined in (3.4) and (3.5) satisfy some certain conditions (similar to the conditions in [\(Douc, Doukhan and Moulines, 2013, Proposition 21\)](#)), then the strong consistency holds.

3.2. Convergence of the MLE. We always assume in this part that Assumption [\(AG-1\)](#) holds throughout. Note that every ODM of any order $p \geq 1$ can be embedded in a first-order ODM. Therefore, the approach used to derive the convergence of the MLE in the class of first-order models can be applied to the class of higher-order ones up to some necessary adaptations. The following is a list of additional assumptions sufficient for obtaining the convergence of the MLE for a class of higher-order ODMs.

- [\(AG-2\)](#) There exists a function $\bar{V} : \mathbf{Z} \rightarrow \mathbb{R}_+$ such that, for all $\theta \in \Theta$, $\pi_1^\theta(\bar{V}) < \infty$.

REMARK 2. Assumption [\(AG-2\)](#) is usually obtained as a byproduct of the proof of Assumption [\(AG-1\)](#); see [Section 3.4](#). It is here stated as an assumption for convenience.

The following set of conditions can readily be checked on g^θ , ψ^θ and Ψ^θ .

- [\(BG-1\)](#) For all $y \in \mathbf{Y}$, the function $(\theta, x) \mapsto g^\theta(x; y)$ is continuous on $\Theta \times \mathbf{X}$.
- [\(BG-2\)](#) For all $y_{1:p} \in \mathbf{Y}$, the function $(\theta, x_{1:p}) \mapsto \psi_{y_{1:p}}^\theta(x_{1:p})$ is continuous on $\Theta \times \mathbf{X}^p$.

The function \bar{V} appearing in [\(BG-3\)\(viii\)](#) below is the same one as in Assumption [\(AG-2\)](#). Moreover, in this condition and in what follows, we write $f \lesssim V$ for a real-valued function f and a nonnegative function V defined

on the same space Z , whenever there exists a positive constant c such that $|f(z)| \leq cV(z)$ for all $z \in Z$.

(BG-3) There exist $z_1 \in Z$, a set $Z_1 \subseteq Z$ such that $\Pi_p(Z_1)$ is closed, $\varrho \in (0, 1)$, $C \geq 0$ and measurable functions $\bar{\Psi} : Z \rightarrow \mathbb{R}_+$, $H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and $\bar{\phi} : Y \rightarrow \mathbb{R}_+$ such that the following assertions hold.

(i) For all $\theta \in \Theta$ and $(z, y) \in Z \times Y$, $\Pi_p(\Psi_y^\theta(z)) \in Z_1$.

(ii) $\sup_{(\theta, z, y) \in \Theta \times Z_1 \times Y} g^\theta(\Pi_p(z); y) < \infty$.

(iii) For all $\theta \in \Theta$, $n \in \mathbb{Z}_+$, $z \in Z$, and $y_{1:n} \in Y^n$,

$$(3.8) \quad d_Z \left(\Psi^\theta \langle y_{1:n} \rangle (z_1), \Psi^\theta \langle y_{1:n} \rangle (z) \right) \leq \varrho^n \bar{\Psi}(z),$$

(iv) $\bar{\Psi}$ is locally bounded.

(v) For all $\theta \in \Theta$ and $y \in Y$, $\bar{\Psi}(\Psi_y^\theta(z_1)) \leq \bar{\phi}(y)$.

(vi) For all $\theta \in \Theta$ and $(z, z', y) \in Z_1 \times Z_1 \times Y$,

$$(3.9) \quad \left| \ln \frac{g^\theta(\Pi_p(z); y)}{g^\theta(\Pi_p(z'); y)} \right| \leq H(d_Z(z, z')) e^{C(d_Z(z_1, z) \vee d_Z(z_1, z'))} \bar{\phi}(y),$$

(vii) $H(u) = O(u)$ as $u \rightarrow 0$.

(viii) If $C = 0$, then, for all $\theta \in \Theta$,

$$(3.10) \quad G^\theta \ln^+ \bar{\phi} \lesssim \bar{V};$$

otherwise, for all $\theta \in \Theta$,

$$(3.11) \quad G^\theta \bar{\phi} \lesssim \bar{V}.$$

Let us now state the equivalence-class consistency of the MLE for the higher-order ODMs. The proof can be adapted and easily follows from (Douc, Roueff and Sim, 2015, Theorem 1) and is thus omitted here.

THEOREM 3. *Assume that (AG-1), (AG-2), (BG-1), (BG-2) and (BG-3) hold. Then, letting $z_1 \in Z$ as in (BG-3), the convergence (3.7) of the MLE holds.*

The *strong* consistency of the MLE follows from the following theorem.

THEOREM 4. *Assume that (AG-1), (AG-2), (BG-1), (BG-2) and (BG-3) hold. Suppose in addition that*

(a) for all $\theta = (\vartheta, r), \theta' = (\vartheta', r') \in \Theta$ and $x, x' \in \mathsf{X}$,

$$G^\theta(x; \cdot) = G^{\theta'}(x'; \cdot) \quad \text{implies} \quad r = r' \quad \text{and} \quad x = x',$$

(b) for all $\theta = (\vartheta, r), \theta_\star = (\vartheta_\star, r_\star) \in \Theta$, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\Pi_p(\Psi^{(\vartheta, r)}\langle Y_{-\infty:0} \rangle) = \Pi_p(\Psi^{(\vartheta_\star, r_\star)}\langle Y_{-\infty:0} \rangle) \quad \text{implies} \quad \vartheta = \vartheta_\star,$$

where $\Psi^\theta\langle Y_{-\infty:0} \rangle$ is defined by (3.4) and (3.5). Then, letting $z_1 \in \mathsf{Z}$ as in (BG-3),

$$(3.12) \quad \lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{z_1, n}, \theta_\star) = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

PROOF. From Theorem 3, we have $\Theta_\star = [\theta_\star]$, where Θ_\star is given in (3.2). Now let $\theta = (\vartheta, r) \in [\theta_\star]$ and write $\theta_\star = (\vartheta_\star, r_\star)$. The Kulback-Leibler divergence implies, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$G^{(\vartheta, r)}\left(\Pi_p(\Psi^{(\vartheta, r)}\langle Y_{-\infty:0} \rangle); \cdot\right) = G^{(\vartheta_\star, r_\star)}\left(\Pi_p(\Psi^{(\vartheta_\star, r_\star)}\langle Y_{-\infty:0} \rangle); \cdot\right).$$

Then from (a), we obtain

$$r = r_\star$$

and

$$\Pi_p(\Psi^{(\vartheta, r)}\langle Y_{-\infty:0} \rangle) = \Pi_p(\Psi^{(\vartheta_\star, r_\star)}\langle Y_{-\infty:0} \rangle), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

From (b), we also have $\vartheta = \vartheta_\star$. Thus the proof follows. \square

3.3. Asymptotic Normality of the MLE. In this section, we treat a special but an important case of ODMs where the space Z is a subset of a finite-dimensional vector space; Θ is a compact subset of \mathbb{R}^d , for some positive integer d ; and the function $\Psi_y^\theta(z)$ defined in (2.6) admits the following form: for all $(z, y, \theta) \in \mathsf{Z} \times \mathsf{Y} \times \Theta$,

$$(3.13) \quad \Psi_y^\theta(z) = a(\theta, y)z + b(\theta, y),$$

where $a(\theta, y)$ and $b(\theta, y)$ are two matrices with appropriate dimensions, and for all $\theta \in \Theta$, $y \mapsto a(\theta, y)$ and $y \mapsto b(\theta, y)$ are measurable maps on $(\mathsf{Y}, \mathcal{Y})$. The underlying model is assumed to be well-specified and $\theta_\star \in \Theta$ denotes the true parameter. As in Section 3.2, Assumption (AG-1) is supposed to hold throughout this section. Assumption (CG-1) below guarantees the existence of $\Psi^\theta\langle \cdot \rangle$ satisfying (3.4) and (3.5), and its differentiability on Θ . In what follows, we denote by $|\cdot|$ a vector norm on \mathbb{R}^d and by $\|\cdot\|$ a matrix norm.

(CG-1) We have the following.

- (i) For all $y \in Y$, $\theta \mapsto a(\theta, y)$ and $\theta \mapsto b(\theta, y)$ are twice continuously differentiable on Θ .
- (ii) There exist constants $C > 0$ and $\rho \in (0, 1)$ such that for all $\theta \in \Theta$ and $y_{1:n} \in Y$,

$$\left\| \prod_{\ell=1}^n a(\theta, y_\ell) \right\| \leq C\rho^n.$$

- (iii) $\tilde{\mathbb{E}}^{\theta_*}[\sup_{\theta \in \Theta} \|b(\theta, Y_1)\|] < \infty$.

- (iv) For all $i \in \{1, \dots, d\}$,

$$\tilde{\mathbb{E}}^{\theta_*} \left[\sup_{\theta \in \Theta} \left\| \frac{\partial a(\theta, Y_1)}{\partial \theta_i} \right\| \right] + \tilde{\mathbb{E}}^{\theta_*} \left[\sup_{\theta \in \Theta} \left\| \frac{\partial b(\theta, Y_1)}{\partial \theta_i} \right\| \right] < \infty.$$

- (v) For all $i, j \in \{1, \dots, d\}$,

$$\tilde{\mathbb{E}}^{\theta_*} \left[\sup_{\theta \in \Theta} \left\| \frac{\partial^2 a(\theta, Y_1)}{\partial \theta_i \partial \theta_j} \right\| \right] + \tilde{\mathbb{E}}^{\theta_*} \left[\sup_{\theta \in \Theta} \left\| \frac{\partial^2 b(\theta, Y_1)}{\partial \theta_i \partial \theta_j} \right\| \right] < \infty.$$

LEMMA 5. Assume (AG-1) and (CG-1). Then for all $\theta \in \Theta$, there exists a $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. finite measurable function $\Psi^\theta \langle \cdot \rangle : Z^{\mathbb{Z}^-} \rightarrow Z$ satisfying (3.4) and (3.5) for all $z \in Z$. The function $\theta \mapsto \Psi^\theta \langle Y_{-\infty:0} \rangle$ is $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. twice continuously differentiable on Θ and uniformly on Θ , we have $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.,

$$(3.14) \quad \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta} = \lim_{m \rightarrow \infty} \frac{\partial \Psi^\theta \langle Y_{-m:0} \rangle(z)}{\partial \theta}$$

$$(3.15) \quad \frac{\partial^2 \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta^T \partial \theta} = \lim_{m \rightarrow \infty} \frac{\partial^2 \Psi^\theta \langle Y_{-m:0} \rangle(z)}{\partial \theta^T \partial \theta}.$$

Moreover,

$$(3.16) \quad \tilde{\mathbb{E}}^{\theta_*} \left[\sup_{\theta \in \Theta} \left\| \Psi^\theta \langle Y_{-\infty:0} \rangle \right\| \right] < \infty,$$

$$(3.17) \quad \tilde{\mathbb{E}}^{\theta_*} \left[\sup_{\theta \in \Theta} \left\| \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta} \right\| \right] < \infty,$$

$$(3.18) \quad \tilde{\mathbb{E}}^{\theta_*} \left[\sup_{\theta \in \Theta} \left\| \frac{\partial^2 \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta^T \partial \theta} \right\| \right] < \infty.$$

PROOF. The proof is postponed to Section 6.1 for convenience. \square

- (CG-2) For all $y \in Y$, the function $(\theta, x) \mapsto g^\theta(x; y)$ twice continuously differentiable on $\Theta \times X$.

For stating further assumptions, the following notation may be needed. For all $k \in \mathbb{Z}_+^*$, $\theta \in \Theta$ and $z \in \mathcal{Z}$, let

$$\ell_{z,k}^\theta := \ln g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{1:k-1} \rangle(z)); Y_k \right).$$

Then for all $n \in \mathbb{Z}_+^*$, it follows that $\mathbf{L}_{z,n}^\theta$ defined by (2.5) can be rewritten as

$$(3.19) \quad \mathbf{L}_{z,n}^\theta = n^{-1} \sum_{k=1}^n \ell_{z,k}^\theta$$

For all $k \in \mathbb{Z}_+^*$, $\theta \in \Theta$ and $z \in \mathcal{Z}$, let

$$\ell_k^\theta := \ln g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:k-1} \rangle); Y_k \right),$$

which is $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. well defined by Lemma 5. Note that from (3.6), we have for all $k \in \mathbb{Z}_+^*$, $\theta \in \Theta$, $\ell_k^\theta = \ln p^\theta(Y_k | Y_{-\infty:k-1})$. Now for all $n \in \mathbb{Z}_+^*$ and $\theta \in \Theta$, define

$$(3.20) \quad \mathbf{L}_n^\theta := n^{-1} \sum_{k=1}^n \ell_k^\theta.$$

REMARK 3. Note that by Lemma 5, (AG-1), (CG-1) and (CG-2) imply that the functions $\theta \mapsto \ell_{z,k}^\theta$ and $\theta \mapsto \ell_k^\theta$ are twice continuously differentiable on some neighborhood $\mathcal{V}(\theta_\star)$ of θ_\star , $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.

Assumptions (AG-1), (CG-1) and (CG-2) together with the following set of assumptions, (CG-3)–(CG-7), are sufficient to yield the asymptotic normality of $\hat{\theta}_{z,n}$ for arbitrary $z \in \mathcal{Z}$.

- (CG-3) The true parameter θ_\star lies within the interior of Θ .
- (CG-4) The strong consistency holds, that is, $\lim_{n \rightarrow \infty} \hat{\theta}_{z,n} = \theta_\star$, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.
- (CG-5) There exist a constant $\alpha > 1/2$ and a $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. finite random variable $C > 0$ such that for all $k \geq 1$,

$$\sup_{\theta \in \mathcal{V}(\theta_\star)} \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta} - \frac{\partial \ell_k^\theta}{\partial \theta} \right| \leq C k^{-\alpha}, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

where $\mathcal{V}(\theta_\star)$ is a neighborhood of θ_\star .

- (CG-6) There exists a neighborhood $\mathcal{V}(\theta_\star)$ of θ_\star such that

$$(i) \quad \tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \mathcal{V}(\theta_\star)} \left| \frac{\partial \ell_1^\theta}{\partial \theta} \right|^2 \right] < \infty.$$

- (ii) $\tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \mathcal{V}(\theta_\star)} \left\| \frac{\partial^2 \ell_1^\theta}{\partial \theta^T \partial \theta} \right\| \right] < \infty.$
- (iii) $\tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \mathcal{V}(\theta_\star)} \left| \frac{\partial}{\partial \theta} g^\theta (\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) \right| \right] < \infty.$
- (iv) $\tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \mathcal{V}(\theta_\star)} \left\| \frac{\partial^2}{\partial \theta^T \partial \theta} g^\theta (\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) \right\| \right] < \infty.$

(CG-7) The matrix $\Sigma_\star = -\tilde{\mathbb{E}}^{\theta_\star} \left[\frac{\partial^2 \ell_1^{\theta_\star}}{\partial \theta^T \partial \theta} \right]$ is invertible.

Note that if Assumptions (AG-1), (CG-1) and (CG-2) hold, then by Remark 3, $\theta \mapsto \mathbf{L}_{z,n}^\theta$ and $\theta \mapsto \mathbf{L}_n^\theta$ respectively given by (3.19) and (3.20) are twice continuously differentiable on some neighborhood $\mathcal{V}(\theta_\star)$ of θ_\star , $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. If in addition Assumption (CG-5) holds, the following fact is obtained.

LEMMA 6. Assume (AG-1), (CG-1), (CG-2) and (CG-5), then

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \mathcal{V}(\theta_\star)} n^{1/2} \left| \frac{\partial \mathbf{L}_{z,n}^\theta}{\partial \theta} - \frac{\partial \mathbf{L}_n^\theta}{\partial \theta} \right| = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

PROOF. We have from (CG-5), $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\sup_{\theta \in \mathcal{V}(\theta_\star)} n^{1/2} \left| \frac{\partial \mathbf{L}_{z,n}^\theta}{\partial \theta} - \frac{\partial \mathbf{L}_n^\theta}{\partial \theta} \right| \leq \frac{1}{n^{1/2}} \sum_{k=1}^n \sup_{\theta \in \mathcal{V}(\theta_\star)} \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta} - \frac{\partial \ell_k^\theta}{\partial \theta} \right| \leq \frac{C}{n^{1/2}} \sum_{k=1}^n \frac{1}{k^\alpha}.$$

To complete, observe that the rightmost term of the above inequalities converges to 0 as $n \rightarrow \infty$ whenever $\alpha > 1/2$. \square

THEOREM 7. Assume that (AG-1) and (CG-1)–(CG-7) hold. Then, for any $z \in \mathcal{Z}$,

$$(3.21) \quad \sqrt{n}(\hat{\theta}_{z,n} - \theta_\star) \rightsquigarrow_{\theta_\star} \mathcal{N}(0, \Sigma_\star^{-1}),$$

where the symbol $\rightsquigarrow_{\theta_\star}$ means weak convergence under $\tilde{\mathbb{P}}^{\theta_\star}$, and $\mathcal{N}(0, \Sigma_\star^{-1})$ stands for the centered Gaussian distribution with covariance matrix Σ_\star^{-1} .

PROOF. From Lemma 6, we have

$$\lim_{n \rightarrow \infty} n^{1/2} \sup_{\theta \in \mathcal{V}(\theta_\star)} \left| \frac{\partial \mathbf{L}_{z,n}^\theta}{\partial \theta} - \frac{\partial \mathbf{L}_n^\theta}{\partial \theta} \right| = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Since for sufficiently large n , $\hat{\theta}_{z,n} \in \mathcal{V}(\theta_\star)$ and $\frac{\partial \hat{\theta}_{z,n}}{\partial \theta} = 0$, then

$$n^{1/2} \frac{\partial \hat{\theta}_{z,n}}{\partial \theta} = n^{1/2} \frac{\partial \hat{\theta}_{z,n}}{\partial \theta} + o_P(1) = o_P(1).$$

Note that we also have,

$$n^{1/2} \frac{\partial \mathbf{L}_n^{\hat{\theta}_{z,n}}}{\partial \theta} = n^{1/2} \frac{\partial \mathbf{L}_n^{\theta_\star}}{\partial \theta} + \left(\frac{\partial^2 \mathbf{L}_n^{\theta_\star}}{\partial \theta^T \partial \theta} + \varepsilon_n \right) n^{1/2} (\hat{\theta}_{z,n} - \theta_\star),$$

where

$$(3.22) \quad \varepsilon_n = \int_0^1 \left(\frac{\partial^2 \mathbf{L}_n^{\theta_\star + t(\hat{\theta}_{z,n} - \theta_\star)}}{\partial \theta^T \partial \theta} - \frac{\partial^2 \mathbf{L}_n^{\theta_\star}}{\partial \theta^T \partial \theta} \right) dt.$$

Thus,

$$n^{1/2} \frac{\partial \mathbf{L}_n^{\theta_\star}}{\partial \theta} = - \left(\frac{\partial^2 \mathbf{L}_n^{\theta_\star}}{\partial \theta^T \partial \theta} + \varepsilon_n \right) n^{1/2} (\hat{\theta}_{z,n} - \theta_\star) + o_P(1).$$

Note that from (CG-6)(ii) and by ergodicity, Birkhoff ergodic theorem applies and we have

$$\lim_{n \rightarrow \infty} \frac{\partial^2 \mathbf{L}_n^{\theta_\star}}{\partial \theta^T \partial \theta} = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \frac{\partial^2 \ell_k^{\theta_\star}}{\partial \theta^T \partial \theta} = -\Sigma_\star, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Thus to complete the proof, we only need to show that

- (a) $n^{1/2} \frac{\partial \mathbf{L}_n^{\theta_\star}}{\partial \theta} \rightsquigarrow_{\theta_\star} \mathcal{N}(0, \Sigma_\star)$ and
- (b) $\varepsilon_n = o_P(1)$.

Let us now we show (a). Note that we can write

$$n^{1/2} \frac{\partial \mathbf{L}_n^{\theta_\star}}{\partial \theta} = n^{-1/2} \sum_{k=1}^n \frac{\partial \ell_k^{\theta_\star}}{\partial \theta}.$$

Let $\xi_k = \frac{\partial \ell_k^{\theta_\star}}{\partial \theta}$, $k \in \mathbb{Z}_+$. From (CG-6)(iii), it readily follows that

$$\begin{aligned} \tilde{\mathbb{E}}^{\theta_\star} [\xi_k | Y_{-\infty:k-1}] &= \int_{\mathcal{Y}} \frac{\partial g^{\theta_\star}(\Pi_p(\Psi^{\theta_\star}(Y_{-\infty:k-1})); y)}{\partial \theta} \nu(dy) \\ &= \frac{\partial}{\partial \theta} \int_{\mathcal{Y}} g^\theta(\Pi_p(\Psi^\theta(Y_{-\infty:k-1})); y) \nu(dy) \Big|_{\theta=\theta_\star} \\ &= \frac{\partial}{\partial \theta} 1 \Big|_{\theta=\theta_\star} = 0, \end{aligned}$$

so that the sequence $\{\xi_k : k \in \mathbb{Z}_+\}$ is a stationary and ergodic Martingale difference with respect to $\mathcal{F}_{-\infty:k} = \sigma(Y_{-\infty:k})$ in $L_2(\tilde{\mathbb{P}}^{\theta_\star})$. By (CG-6) and (CG-7), its covariance matrix is equal to Σ_\star . To see this, observe that $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\begin{aligned} & \tilde{\mathbb{E}}^{\theta_\star} [\xi_k \xi_k^T | \mathcal{F}_{-\infty:k-1}] \\ &= \int_{\mathcal{Y}} \frac{1}{g^{\theta_\star}(\psi_k^{\theta_\star}; y)} \left(\frac{\partial g^{\theta_\star}(\psi_k^{\theta_\star}; y)}{\partial \theta} \right) \left(\frac{\partial g^{\theta_\star}(\psi_k^{\theta_\star}; y)}{\partial \theta} \right)^T \nu(dy) \\ &= \int_{\mathcal{Y}} \left(\frac{\partial^2 g^{\theta_\star}(\psi_k^{\theta_\star}; y)}{\partial \theta^T \partial \theta} - \frac{\partial^2 \ln g^{\theta_\star}(\psi_k^{\theta_\star}; y)}{\partial \theta^T \partial \theta} g^{\theta_\star}(\psi_k^{\theta_\star}; y) \right) \nu(dy), \end{aligned}$$

with $\psi_k^{\theta_\star} := \Pi_p(\Psi^{\theta_\star} \langle Y_{-\infty:k-1} \rangle)$. By (CG-6)(iv), we have

$$\int_{\mathcal{Y}} \frac{\partial^2 g^{\theta_\star}(\psi_k^{\theta_\star}; y)}{\partial \theta^T \partial \theta} = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Since $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\tilde{\mathbb{E}}^{\theta_\star} \left[\frac{\partial^2 \ell_k^{\theta_\star}}{\partial \theta^T \partial \theta} \middle| \mathcal{F}_{-\infty:k-1} \right] = \int_{\mathcal{Y}} \frac{\partial^2 \ln g^{\theta_\star}(\psi_k^{\theta_\star}; y)}{\partial \theta^T \partial \theta} g^{\theta_\star}(\psi_k^{\theta_\star}; y) \nu(dy),$$

it then follows that

$$\tilde{\mathbb{E}}^{\theta_\star} [\xi_k \xi_k^T | \mathcal{F}_{-\infty:k-1}] = -\tilde{\mathbb{E}}^{\theta_\star} \left[\frac{\partial^2 \ell_k^{\theta_\star}}{\partial \theta^T \partial \theta} \middle| \mathcal{F}_{-\infty:k-1} \right], \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Taking the expectation on both sides, then using tower property and using (CG-7), the claim thus follows. By ergodicity and from (CG-6)(i), we have

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \tilde{\mathbb{E}}^{\theta_\star} [\xi_k \xi_k^T | \mathcal{F}_{-\infty:k-1}] = \Sigma_\star, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Moreover, for $\epsilon > 0$,

$$\begin{aligned}
& \lim_{n \rightarrow \infty} n^{-1} \left\| \sum_{k=1}^n \tilde{\mathbb{E}}^{\theta_*} \left[\xi_k \xi_k^T \mathbb{1}(|\xi_k| > \epsilon \sqrt{n}) \middle| \mathcal{F}_{-\infty:k-1} \right] \right\| \\
& \leq \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \tilde{\mathbb{E}}^{\theta_*} \left[|\xi_k|^2 \mathbb{1}(|\xi_k| > \epsilon \sqrt{n}) \middle| \mathcal{F}_{-\infty:k-1} \right] \\
& \leq \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \tilde{\mathbb{E}}^{\theta_*} \left[|\xi_k|^2 \mathbb{1}(|\xi_k| > M) \middle| \mathcal{F}_{-\infty:k-1} \right] \\
& = \tilde{\mathbb{E}}^{\theta_*} \left[|\xi_1|^2 \mathbb{1}(|\xi_1| > M) \right], \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.},
\end{aligned}$$

where $M > 0$. Since M is arbitrary, letting $M \rightarrow \infty$, we have

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \tilde{\mathbb{E}}^{\theta_*} \left[\xi_k \xi_k^T \mathbb{1}(|\xi_k| > \epsilon \sqrt{n}) \middle| \mathcal{F}_{-\infty:k-1} \right] = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

Thus by martingale central limit theorem, see, for instance, [Hall and Heyde \(1980\)](#),

$$n^{1/2} \frac{\partial \mathbf{L}_n^{\theta_*}}{\partial \theta} = n^{-1/2} \sum_{k=1}^n \xi_k \rightsquigarrow_{\theta_*} \mathcal{N}(0, \Sigma_*).$$

Now it remains to show (b). Let $\rho > 0$ such that $\mathbf{B}(\theta_*, \rho) \subset \mathcal{V}(\theta_*)$. Since $\tilde{\mathbb{P}}^{\theta_*}$ -a.s., $\hat{\theta}_{z,n} \rightarrow \theta_*$ as $n \rightarrow \infty$, then there exists a $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. finite integer N_0 such that for all $n \geq N_0$, $\tilde{\mathbb{P}}^{\theta_*}$ -a.s., $\theta_* + t(\hat{\theta}_{z,n} - \theta_*) \in \mathbf{B}(\theta_*, \rho)$. Then from (3.22), $\tilde{\mathbb{P}}^{\theta_*}$ -a.s., for all $n \geq N_0$,

$$\begin{aligned}
\|\varepsilon_n\| & \leq \int_0^1 \left\| \frac{\partial^2 \mathbf{L}_n^{\theta_* + t(\hat{\theta}_{z,n} - \theta_*)}}{\partial \theta^T \partial \theta} - \frac{\partial^2 \mathbf{L}_n^{\theta_*}}{\partial \theta^T \partial \theta} \right\| dt \\
& \leq \sup_{\theta \in \mathbf{B}(\theta_*, \rho)} \left\| \frac{\partial^2 \mathbf{L}_n^\theta}{\partial \theta^T \partial \theta} - \frac{\partial^2 \mathbf{L}_n^{\theta_*}}{\partial \theta^T \partial \theta} \right\| \\
& \leq n^{-1} \sum_{k=1}^n \sup_{\theta \in \mathbf{B}(\theta_*, \rho)} \left\| \frac{\partial^2 \ell_k^\theta}{\partial \theta^T \partial \theta} - \frac{\partial^2 \ell_k^{\theta_*}}{\partial \theta^T \partial \theta} \right\|.
\end{aligned}$$

By (CG-6)(ii), $\mathbf{B}(\theta_*, \rho) \subset \mathcal{V}(\theta_*)$ and Birkhoff ergodic theorem, we get that

$$\limsup_{n \rightarrow \infty} \|\varepsilon_n\| \leq \tilde{\mathbb{E}}^{\theta_*} \left[\sup_{\theta \in \mathbf{B}(\theta_*, \rho)} \left\| \frac{\partial^2 \ell_1^\theta}{\partial \theta^T \partial \theta} - \frac{\partial^2 \ell_1^{\theta_*}}{\partial \theta^T \partial \theta} \right\| \right], \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

Here ρ is an arbitrary positive number, provided that $\mathbf{B}(\theta_*, \rho) \subset \mathcal{V}(\theta_*)$. By dominated convergence and [Remark 3](#), the right-hand side of the previous display can be made arbitrarily close to zero by letting ρ tend to zero. Hence we get [\(b\)](#), and the proof is complete. \square

REMARK 4. From the proof of [Theorem 7](#), it is also shown that

$$\Sigma_\star = \tilde{\mathbb{E}}^{\theta_\star} \left[\begin{pmatrix} \frac{\partial \ell_1^{\theta_\star}}{\partial \theta} \\ \left(\frac{\partial \ell_1^{\theta_\star}}{\partial \theta} \right)^T \end{pmatrix} \right].$$

Thus, to obtain [\(CG-7\)](#), that is showing that Σ_\star is invertible, it is sufficient to show that for all $\mathbf{a} \in \mathbb{R}^d$,

$$(3.23) \quad \mathbf{a}^T \frac{\partial \ell_1^{\theta_\star}}{\partial \theta} = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}, \implies \mathbf{a} = 0.$$

3.4. *Ergodicity.* In this section, we provide conditions that yield stationarity and ergodicity of the Markov chain $\{(Z_k, Y_k) : k \in \mathbb{Z}_+\}$. We will set θ to be an arbitrary value in Θ and since this is a “for all θ (...)” condition, to save space and alleviate the notational burden, we will drop the superscript θ from, for example, G^θ , R^θ and ψ^θ and respectively write G , R and ψ , instead.

Ergodicity of Markov chains are usually studied using φ -irreducibility. This approach is well known to be quite efficient when dealing with fully dominated models; see [Meyn and Tweedie \(2009\)](#). It is not at all the same picture for integer-valued observation-driven models, where other tools need to be invoked; see [Fokianos and Tjøstheim \(2011\)](#); [Douc, Doukhan and Moulines \(2013\)](#); [Douc, Roueff and Sim \(2015\)](#). Our result here is obtained in the same fashion as in [Douc, Roueff and Sim \(2015\)](#), which is inspired by the approach in [Douc, Doukhan and Moulines \(2013\)](#). The following is the list of assumptions to obtain our main result on ergodicity. We should mention that some of our assumptions (for instance, Assumption [\(AG-4\)](#)) are more general than those derived in [Douc, Doukhan and Moulines \(2013\)](#) and [Douc, Roueff and Sim \(2015\)](#).

- (AG-3) The measurable space (Z, d_Z) is a locally compact, complete and separable metric space and its associated σ -field \mathcal{Z} is the Borel σ -field.
- (AG-4) There exists a positive integer q such that the Markov kernel R^q is weak Feller, that is, for any continuous and bounded function f defined on X , $R^q f$ is continuous and bounded on X . Moreover, there exist $(\lambda, \beta) \in (0, 1) \times \mathbb{R}_+$ and a measurable function $V : Z \rightarrow \mathbb{R}_+$ such that $R^q V \leq \lambda V + \beta$ and $\{V \leq M\}$ is a compact set for any $M > 0$.

(AG-5) The Markov kernel R admits a reachable point, that is, there exists $z_0 \in Z$ such that, for any $z \in Z$ and any neighborhood \mathcal{N} of z_0 , $R^m(z; \mathcal{N}) > 0$ for at least one positive integer m .

(AG-6) There exists a positive integer ℓ such that

$$(3.24) \quad \sup_{\substack{(z, z', y_{1:\ell}) \in Z^2 \times Y^\ell \\ z \neq z'}} \frac{d_Z(\Psi\langle y_{1:\ell} \rangle(z), \Psi\langle y_{1:\ell} \rangle(z'))}{d_Z(z, z')} < 1,$$

and

$$(3.25) \quad \sup_{\substack{(z, z', y) \in Z^2 \times Y \\ z \neq z'}} \frac{d_Z(\Psi_y(z), \Psi_y(z'))}{d_Z(z, z')} < \infty.$$

(AG-7) There exist a measurable function α from Z^2 to $[0, 1]$, a measurable function $\phi : Z^2 \rightarrow X = \Pi_p(Z)$ and a measurable function $W : Z^2 \rightarrow [1, \infty)$ such that the following assertions hold.

(i) For all $(z, z') \in Z^2$ and $y \in Y$,

$$(3.26) \quad \min \{g(\Pi_p(z); y), g(\Pi_p(z'); y)\} \geq \alpha(z, z')g(\phi(z, z'); y).$$

(ii) For all $z \in Z$, $W(z, \cdot)$ is finitely bounded in a neighborhood of z , that is, there exists $\gamma_z > 0$ such that $\sup_{z' \in B(z, \gamma_z)} W(z, z') < \infty$.

(iii) For all $(z, z') \in Z^2$, $1 - \alpha(z, z') \leq d_Z(z, z')W(z, z')$.

(iv) There exist an integer $m > 0$ and a real number $D > 0$ such that

$$\sup_{(z, z') \in Z^2} (J_m((z, z'); W) - W(z, z')) < \infty$$

and

$$\sup_{(z, z') \in Z^2} (J_1((z, z'); W) - DW(z, z')) < \infty,$$

where for any positive integer n , $(z, z') \in Z^2$ and measurable function $f : Z^2 \rightarrow \mathbb{R}$,

$$J_n((z, z'); f) := \int f(\Psi\langle y_{1:n} \rangle(z), \Psi\langle y_{1:n} \rangle(z')) \times \prod_{k=1}^n G(\phi(\Psi\langle y_{1:k-1} \rangle(z), \Psi\langle y_{1:k-1} \rangle(z'))); dy_k).$$

REMARK 5. Assumption (AG-3) holds if the measurable metric spaces (X, d_X) and (Y, d_Y) are locally compact, complete and separable and their associated σ -field \mathcal{X} and \mathcal{Y} are the Borel σ -fields. In this case, the metric d_Z can be appropriately defined as a function of d_X and d_Y .

REMARK 6. Note that if the Markov kernel R is weak Feller, then the Markov kernel R^q is also weak Feller.

LEMMA 8. Assume (AG-3) and (AG-4). Then R admits an invariant probability distribution $\tilde{\pi}$; moreover, $\tilde{\pi}V < \infty$.

PROOF. From Assumptions (AG-3) and (AG-4) and by Remark 6, the transition kernel R^q admits an invariant probability distribution denoted by π_q . Let $\tilde{\pi}$ be defined by, for all $A \in \mathcal{Z}$,

$$\tilde{\pi}(A) = \frac{1}{q} \sum_{k=1}^q \pi_q R^k(A).$$

Obviously, we have $\tilde{\pi}R = \tilde{\pi}$, which shows that R admits an invariant probability distribution $\tilde{\pi}$. Now let $M > 0$. Then we have for all $n \in \mathbb{Z}_+$,

$$\begin{aligned} \tilde{\pi}(V \wedge M) &= \tilde{\pi}R^{nq}(V \wedge M) \\ &\leq \tilde{\pi}R^{nq}(V) \wedge M \\ &\leq \lambda^n \tilde{\pi}(V) \wedge M + \frac{\beta}{1-\lambda} \wedge M. \end{aligned}$$

Letting $n \rightarrow \infty$, we then obtain

$$\tilde{\pi}(V \wedge M) \leq \frac{\beta}{1-\lambda} \wedge M.$$

Finally, since M is arbitrary, by letting $M \rightarrow \infty$ and by monotone convergence theorem, we get $\tilde{\pi}V < \infty$ as desired. \square

We can now state the main ergodicity result.

THEOREM 9. Conditions (AG-3), (AG-4), (AG-5), (AG-6) and (AG-7) imply that K admits a unique stationary distribution π on $Z \times Y$. Moreover, $\pi_1 \bar{V} < \infty$ for every $\bar{V} : X \rightarrow \mathbb{R}_+$ such that $\bar{V} \lesssim V$.

For convenience, we postponed the proof of Theorem 9 to Section 6.2. The first conclusion of Theorem 9 can directly be applied for all $\theta \in \Theta$ to check (AG-1).

Assumptions (AG-4) and (AG-5) have to be checked directly on the Markov kernel R defined by (2.11). To this end, it can be useful to define, for any given $z \in \mathbf{Z}$, the distribution

$$(3.27) \quad \bar{\mathbb{P}}_z := \mathbb{P}_{\delta_z \otimes G(\Pi_p(z); \cdot)}$$

on $(\mathbf{Z} \times \mathbf{Y})^{\mathbb{Z}_+}$, where \mathbb{P}_ξ is defined for any distribution ξ on $\mathbf{Z} \times \mathbf{Y}$ as in Definition 2. Then the first component process $\{Z_k : k \in \mathbb{Z}_+\}$ associated to $\bar{\mathbb{P}}_z$ is a Markov chain with Markov kernel R and initial distribution δ_z .

The following lemma provides a general way for constructing the instrumental functions α and ϕ that appear in (AG-7). The proof can be easily adapted from (Douc, Roueff and Sim, 2015, Lemma 1) and is thus omitted.

LEMMA 10. *Suppose that $\Pi_p(\mathbf{Z}) = \mathbf{X} = \mathbf{C}^{\mathbf{S}}$ for some measurable space $(\mathbf{S}, \mathcal{S})$ and $\mathbf{C} \subseteq \mathbb{R}$. Thus for all $x \in \mathbf{X}$, we write $x = (x_s)_{s \in \mathbf{S}}$, where $x_s \in \mathbf{C}$ for all $s \in \mathbf{S}$. Suppose moreover that for all $x = (x_s)_{s \in \mathbf{S}} \in \mathbf{X}$, we can express the conditional density $g(x; \cdot)$ as a mixture of densities of the form $j(x_s)h(x_s; \cdot)$ over $s \in \mathbf{S}$. This means that for all $t \in \mathbf{C}$, $y \mapsto j(t)h(t; y)$ is a density with respect to ν and there exists a probability measure μ on $(\mathbf{S}, \mathcal{S})$ such that*

$$(3.28) \quad g(x; y) = \int_{\mathbf{S}} j(x_s)h(x_s; y)\mu(ds), \quad y \in \mathbf{Y}.$$

We moreover assume that h takes nonnegative values and that one of the two following assumptions holds.

- (H'-1) For all $y \in \mathbf{Y}$, the function $h(\cdot; y) : t \mapsto h(t; y)$ is nondecreasing.
- (H'-2) For all $y \in \mathbf{Y}$, the function $h(\cdot; y) : t \mapsto h(t; y)$ is nonincreasing.

For all $(x, x') \in \mathbf{X}^2$, we denote $x \wedge x' := (\min\{x_s, x'_s\})_{s \in \mathbf{S}}$ and $x \vee x' := (\max\{x_s, x'_s\})_{s \in \mathbf{S}}$ and we define $\alpha(z, z')$ and $\phi(z, z')$ by: for all $(z, z') \in \mathbf{Z}^2$, letting $x = \Pi_p(z)$ and $x' = \Pi_p(z')$,

$$\begin{cases} \alpha(z, z') = \inf_{s \in \mathbf{S}} \left\{ \frac{j(x_s \vee x'_s)}{j(x_s \wedge x'_s)} \right\} & \text{and } \phi(z, z') = x \wedge x' & \text{under (H'-1);} \\ \alpha(z, z') = \inf_{s \in \mathbf{S}} \left\{ \frac{j(x_s \wedge x'_s)}{j(x_s \vee x'_s)} \right\} & \text{and } \phi(z, z') = x \vee x' & \text{under (H'-2).} \end{cases}$$

Then α and ϕ defined above satisfy (AG-7)(i).

4. Examples. In this section, we apply our main results derived above to two models of interest, namely, the log-linear Poisson GARCH(p, p) and the NBIN-GARCH(p, p) models. To the best of our knowledge, the stationarity and ergodicity as well as the asymptotic properties of the MLE for

the general log-linear Poisson GARCH(p, p) and NBIN-GARCH(p, p) models have not been derived so far. In the sequel, \mathbf{A} , \mathbf{b} and $\boldsymbol{\omega}$ always denote, respectively, the matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ a_p & a_{p-1} & a_{p-2} & a_{p-3} & \cdots & a_1 & b_p & b_{p-1} & \cdots & b_3 & b_2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

the vector $b_1\boldsymbol{\delta}_p + \boldsymbol{\delta}_{2p-1}$, with for all $j \in \{1, \dots, 2p-1\}$, $\boldsymbol{\delta}_j = (\delta_{ij})_{i \in \{1, \dots, 2p-1\}}$, δ_{ij} being the Kronecker delta, and the vector $\boldsymbol{\omega}\boldsymbol{\delta}_p$. We also denote by \mathbf{A}_* , \mathbf{b}_* and $\boldsymbol{\omega}_*$ the corresponding values of \mathbf{A} , \mathbf{b} and $\boldsymbol{\omega}$ at $(\boldsymbol{\omega}, a_{1:p}, b_{1:p}) = (\boldsymbol{\omega}_*, a_{1:p}^*, b_{1:p}^*)$, respectively. For any matrix \mathbf{M} , we denote its transpose by \mathbf{M}^T . For any real numbers a and b , we denote their maximum value by $a \vee b$ or $\max\{a, b\}$ and their minimum value by $a \wedge b$ or $\min\{a, b\}$.

Now, let \mathbf{I}_n be the identity matrix of order n and denote the characteristic polynomial of the matrix $\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T$ by $P(\lambda) = \det(\lambda\mathbf{I}_{2p-1} - (\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T))$. For any square matrix \mathbf{M} , denote the spectral radius of \mathbf{M} by $|\lambda|_{\max}(\mathbf{M})$. Straightforward computation shows that

$$(4.1) \quad P(\lambda) = \lambda^{p-1} \left(\lambda^p - \sum_{k=1}^p (a_k + b_k) \lambda^{p-k} \right).$$

The following lemma which is useful for the next example is immediate.

LEMMA 11. *Let $a_{1:p} \geq 0$, $b_{1:p} \geq 0$. Then $\sum_{i=1}^p (a_i + b_i) < 1$ if and only if*

$$|\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T) < 1.$$

PROOF. The proof is postponed to [Section 6.3](#). \square

REMARK 7. It is well known that for any nonnegative matrices $\mathbf{M} = (m_{ij})$ and $\mathbf{N} = (n_{ij})$ such that $\mathbf{M} \geq \mathbf{N}$, that is for all i, j , $m_{ij} \geq n_{ij}$, we have $|\lambda|_{\max}(\mathbf{M}) \geq |\lambda|_{\max}(\mathbf{N})$. Thus for our matrices \mathbf{A} and $\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T$, we have $|\lambda|_{\max}(\mathbf{A}) \leq |\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T)$.

4.1. *Log-linear Poisson GARCH(p, p) Model.*

EXAMPLE 1 (Definition of the log linear Poisson GARCH(p, p)). Consider the log-linear Poisson GARCH of order (p, p) , $p \in \mathbb{Z}_+^*$, which is defined as follows: for all $k \in \mathbb{Z}_+$,

$$(4.2) \quad \begin{aligned} X_{k+1} &= \omega + \sum_{i=1}^p a_i X_{k-i+1} + \sum_{i=1}^p b_i \ln(1 + Y_{k-i+1}), \\ Y_{k+1} | \mathcal{F}_k &\sim \mathcal{P}(e^{X_{k+1}}), \end{aligned}$$

where

- the parameter space

$$\theta = (\omega, a_{1:p}, b_{1:p}) \in \Theta \subset \mathbb{R}^{2p+1},$$

- $\mathcal{F}_k = \sigma(X_{-p+1:k+1}, Y_{-p+1:k})$,
- $\mathbf{X} = \mathbb{R}$, $\mathcal{X} = \mathcal{B}(\mathbf{X})$, $\mathbf{Y} = \mathbb{Z}_+$, $\mathcal{Y} = \mathcal{B}(\mathbf{Y})$ and the metrics on \mathbf{X} and \mathbf{Y} are the same usual metric on \mathbb{R} .

By setting for all k , $U_k = \ln(1 + Y_k)$ (hence $\mathbf{U} = \ln(1 + \mathbb{Z}_+)$, $\mathcal{U} = \mathcal{B}(\mathbf{U})$ and $d_{\mathbf{U}}$ is the usual metric on \mathbb{R}), then Model (4.2) can be embedded in Model (4.3) below:

$$(4.3) \quad \begin{aligned} X_{k+1} &= \psi_{U_{k-p+1:k}}^{\theta}(X_{k-p+1:k}), \\ U_{k+1} | \mathcal{F}_k &\sim G^{\theta}(X_{k+1}; \cdot), \end{aligned}$$

where for all $(x_{1:p}, u_{1:p}) \in \mathbf{X}^p \times \mathbf{U}^p$,

$$(4.4) \quad \psi_{u_{k-p+1:k}}^{\theta}(x_{k-p+1:k}) := \omega + \sum_{i=1}^p a_i x_{p-i+1} + \sum_{i=1}^p b_i u_{p-i+1}$$

and $G^{\theta}(x; \cdot)$ is the conditional law of U_k given $X_k = x$ defined by

$$(4.5) \quad G^{\theta}(x; A) = \sum_{u \in A} e^{-e^x} \frac{e^{x(e^u - 1)}}{(e^u - 1)!}, \quad A \in \mathcal{U}.$$

The density g^{θ} of U_k conditionally on $X_k = x$ with respect to the counting measure on \mathbf{U} is given by, for all $u \in \mathbf{U}$,

$$(4.6) \quad g^{\theta}(x; u) = e^{-e^x} \frac{e^{x(e^u - 1)}}{(e^u - 1)!} > 0.$$

By [Definition 1](#), Model [\(4.3\)](#) is an $\text{ODM}(p, p)$ and is dominated by the counting measure on \mathbf{U} . Now let $\mathbf{Z} = \mathbf{X}^p \times \mathbf{U}^{p-1}$ and $\mathcal{Z} = \mathcal{X}^{\otimes p} \otimes \mathcal{U}^{\otimes(p-1)}$ and define the measurable function Ψ^θ similarly as in [\(2.6\)](#). The function Ψ^θ then take a simple linear form

$$(4.7) \quad \Psi_u^\theta(z) = \boldsymbol{\omega} + \mathbf{A}z + u\mathbf{b}, \quad (z, u) \in \mathbf{Z} \times \mathbf{U}.$$

Moreover, denoting $Z_k = (X_{k-p+1:k}, U_{k-p+1:k-1})$ for all k , the process $\{Z_k : k \in \mathbb{Z}_+\}$ is a Markov chain on $(\mathbf{Z}, \mathcal{Z})$ whose transition kernel R^θ is given by [\(2.11\)](#) with G^θ and Ψ^θ defined in [\(4.5\)](#) and [\(4.7\)](#), respectively.

In this section, we will use the vector and the matrix norms defined respectively on \mathbb{R}^{2p-1} and the space of $(2p-1) \times (2p-1)$ matrices as follows:

$$|z|_\infty = \bigvee_{i=1}^{2p-1} |\boldsymbol{\delta}_i^T z| = \bigvee_{i=1}^{2p-1} |z_i|,$$

$$\|\mathbf{M}\|_\infty = \bigvee_{i=1}^{2p-1} \sum_{j=1}^{2p-1} |\mathbf{M}_{i,j}| = \sup_{|z|_\infty \leq 1} |\mathbf{M}z|_\infty.$$

For notational convenience, we will also denote $\mathbf{A}_1 = \mathbf{A}$ and $\mathbf{A}_2 = \mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T$.

PROPOSITION 12. *Suppose that $\theta \in \Theta$ satisfies*

$$(4.8) \quad \bigvee_{(i_1, \dots, i_p) \in \{1, 2\}^p} \left\| \prod_{\ell=1}^p \mathbf{A}_{i_\ell} \right\|_\infty < 1.$$

Then we have the following.

- (i) [Theorem 9](#) holds with the function $\bar{V} : \mathbf{Z} \rightarrow \mathbb{R}_+$ defined by, for all $z \in \mathbf{Z}$,

$$(4.9) \quad \bar{V}(z) = e^{\tau|z|_\infty}, \quad \tau > 0.$$

- (ii) [Theorem 3](#) holds with any $z_1 \in \mathbf{Z}$. If the true parameter $\theta_\star = (\boldsymbol{\omega}^\star, \mathbf{a}_{1:p}^\star, \mathbf{b}_{1:p}^\star)$ moreover satisfies, for all $k \in \mathbb{Z}_+$,

$$(4.10) \quad (\mathbf{A}^k - \mathbf{A}_\star^k)\mathbf{b}_\star = 0 \quad \text{implies} \quad \mathbf{A} = \mathbf{A}_\star,$$

then [Theorem 4](#) also holds.

- (iii) *If the true parameter θ_\star lies in the interior of Θ and satisfies [\(4.10\)](#), then [Theorem 7](#) holds.*

REMARK 8. When $p = 1$, then $\mathbf{A}_1 = \mathbf{A}$ and $\mathbf{A}_2 = \mathbf{A} + \mathbf{b}\delta_p^T$ correspond to a_1 and $a_1 + b_1$, respectively. Condition (4.8) then reduces to

$$|a_1| \vee |a_1 + b_1| < 1.$$

This condition is weaker than the one derived in Douc, Doukhan and Moulines (2013) where $|b_1| < 1$ is also imposed.

REMARK 9. To check (4.10), it suffices to check that there exists $k \in \mathbb{Z}_+$ such that the determinant

$$\det \left(\mathbf{A}_*^k \mathbf{b}_*, \mathbf{A}_*^{k+1} \mathbf{b}_*, \dots, \mathbf{A}_*^{k+2p-2} \mathbf{b}_* \right) \neq 0.$$

REMARK 10. When $p = 1$, (4.10) holds if and only if $b^* \neq 0$.

Before proving Proposition 12, let us show the following fact.

LEMMA 13. *If $Y \sim \mathcal{P}(\zeta)$ with $\zeta > 0$, then for any $\vartheta \in \mathbb{R}$, there exist constants $c_1 = c_1(\vartheta), c_2 = c_2(\vartheta) > 0$ such that*

$$(4.11) \quad \mathbb{E}[(1 + Y)^\vartheta] \leq c_1 + c_2 \zeta^\vartheta \mathbb{1}\{\vartheta > 0\}.$$

PROOF. The proof is postponed to Section 6.4. □

PROOF. **Proof of Proposition 12–(i).** Now set $V(z) = \bar{V}(z) = e^{\tau|z|_\infty}$, $z \in \mathbf{Z}$. From Theorem 9, we need to show that (AG-3), (AG-4), (AG-5), (AG-6) and (AG-7) hold. Assumption (AG-3) holds with the metric $d_{\mathbf{Z}}$ defined by, for all $(z, z') \in \mathbf{Z}^2$,

$$(4.12) \quad d_{\mathbf{Z}}(z, z') = \bigvee_{i=1}^{2p-1} |z_i - z'_i| = |z - z'|_\infty.$$

The fact that R is weak Feller follows from the continuity of the density g with respect to the x component in (4.6) and the continuity of $\Pi_p(\cdot)$. We will show that, with V defined in (4.9), then we have $R^p V \leq \lambda V + \beta$ for some $(\lambda, \beta) \in (0, 1) \times \mathbb{R}_+$. Note that Lemma 13 implies that, for all $\vartheta \in \mathbb{R}$ and $\zeta > 0$,

$$(4.13) \quad \int e^{\vartheta u} G(\zeta; du) \leq c_1 + c_2 e^{\vartheta \zeta} \mathbb{1}\{\vartheta > 0\},$$

where c_1 and c_2 are some positive constants. Note further that for all $z \in \mathbf{Z}$ and $\tau > 0$, we have

$$(4.14) \quad V(z) = e^{\tau|z|_\infty} \leq \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} e^{\sigma\tau z_k} \leq 2(2p-1)e^{\tau|z|_\infty}$$

Without loss of generality, let us assume for the moment that $\tau = 1$. Now, using (4.13) and (4.14) and noting that, for all $z \in \mathbf{Z}$, $z_k = \boldsymbol{\delta}_k^T z$, $\Pi_p(z) = \boldsymbol{\delta}_p^T z$ and $\mathbb{1}\{\sigma \boldsymbol{\delta}_k^T \mathbf{b} > 0\} \leq 1$, we have

$$\begin{aligned} RV(z) &\leq \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} \int e^{\sigma \boldsymbol{\delta}_k^T (\omega + \mathbf{A}z + u\mathbf{b})} G(\boldsymbol{\delta}_p^T z; du) \\ &\leq \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} e^{\sigma \boldsymbol{\delta}_k^T (\omega + \mathbf{A}z)} \int e^{\sigma \boldsymbol{\delta}_k^T \mathbf{b}u} G(\boldsymbol{\delta}_p^T z; du) \\ &\leq \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} e^{\sigma \boldsymbol{\delta}_k^T (\omega + \mathbf{A}z)} \left(\gamma_1 + \gamma_2 e^{\sigma \boldsymbol{\delta}_k^T \mathbf{b} \boldsymbol{\delta}_p^T z} \right), \end{aligned}$$

where γ_1 and γ_2 are some positive constants. This further yields that, with $\mathbf{A}_1 = \mathbf{A}$ and $\mathbf{A}_2 = \mathbf{A} + \mathbf{b} \boldsymbol{\delta}_p^T$,

$$\begin{aligned} RV(z) &\leq \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} \left(\beta_1 e^{\sigma \boldsymbol{\delta}_k^T \mathbf{A}z} + \beta_2 e^{\sigma \boldsymbol{\delta}_k^T (\mathbf{A} + \mathbf{b} \boldsymbol{\delta}_p^T)z} \right) \\ &\leq \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} \left(\beta_1 e^{\sigma \boldsymbol{\delta}_k^T \mathbf{A}_1 z} + \beta_2 e^{\sigma \boldsymbol{\delta}_k^T \mathbf{A}_2 z} \right), \end{aligned}$$

where β_1 and β_2 are some positive constants. By iteration, $R^{m+1}V = R(R^m V)$, we obtain that, for all $m \in \mathbb{Z}_+^*$,

(4.15)

$$\begin{aligned}
R^m V(z) &\leq \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} \sum_{(i_1, \dots, i_p) \in \{1,2\}^m} \beta_{i_1, \dots, i_m} e^{\sigma \delta_k^T (\prod_{\ell=1}^m \mathbf{A}_{i_\ell}) z} \\
&\leq \sum_{(i_1, \dots, i_m) \in \{1,2\}^m} \beta_{i_1, \dots, i_m} \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} e^{\sigma \delta_k^T (\prod_{\ell=1}^m \mathbf{A}_{i_\ell}) z} \\
&\leq \sum_{(i_1, \dots, i_m) \in \{1,2\}^m} \beta_{i_1, \dots, i_m} \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} e^{\|\prod_{\ell=1}^m \mathbf{A}_{i_\ell}\|_\infty |z|} \\
&\leq 2(2p-1) \sum_{(i_1, \dots, i_m) \in \{1,2\}^m} \beta_{i_1, \dots, i_m} e^{\|\prod_{\ell=1}^m \mathbf{A}_{i_\ell}\|_\infty |z|} \\
&\leq \left(2(2p-1) \sum_{i_1, \dots, i_m \in \{1,2\}} \beta_{i_1, \dots, i_m} \right) e^{\vee_{i_1, \dots, i_m \in \{1,2\}} \|\prod_{\ell=1}^m \mathbf{A}_{i_\ell}\|_\infty |z|},
\end{aligned}$$

where, for all $(i_1, \dots, i_m) \in \{1, 2\}^m$, β_{i_1, \dots, i_m} are positive constants. Thus, for all $z \in \mathbb{Z}$, with $m = p$, we get

$$R^p V(z) \leq \gamma_p e^{\eta_p |z|},$$

where

$$\gamma_p := 2(2p-1) \sum_{(i_1, \dots, i_p) \in \{1,2\}^p} \beta_{i_1, \dots, i_p} > 0$$

and

$$\eta_p := \bigvee_{(i_1, \dots, i_p) \in \{1,2\}^p} \left\| \prod_{\ell=1}^p \mathbf{A}_{i_\ell} \right\|_\infty < 1.$$

Since

$$\lim_{|z|_\infty \rightarrow \infty} \frac{e^{\eta_p |z|}}{V(z)} = 0$$

and for all $M > 0$,

$$\sup_{|z|_\infty \leq M} R^p V(z) < \infty,$$

then there exists a pair $(\lambda, \beta) \in (0, 1) \times \mathbb{R}_+$ such that $R^p V \leq \lambda V + \beta$, and hence (AG-4) holds. Note that when $m = p - 1$ and $\tau = 1$, we have $e^{|\Pi_p(z)|} \leq \gamma_{p-1} R^{p-1} V(z)$ for some $\gamma_{p-1} > 0$. We now turn to prove (AG-5).

Note that from (4.8), there exist $\delta \in (0, 1)$ and $c > 0$ such that, for any $k \geq 1$,

$$(4.16) \quad \left\| \mathbf{A}^k \right\|_{\infty} \leq c \delta^k.$$

Hence, $(\mathbf{I}_{2p-1} - \mathbf{A})^{-1} = \mathbf{I}_{2p-1} + \sum_{k \geq 1} \mathbf{A}^k$ is well defined. Now, set $z_{\infty} = (\mathbf{I}_{2p-1} - \mathbf{A})^{-1} \boldsymbol{\omega}$. Let \mathcal{N}_{∞} be an open neighborhood of z_{∞} and let $z \in \mathbf{Z}$. Define recursively the sequence $\{z_k : k \in \mathbb{Z}_+\}$ with $z_0 = z$ and for all $k \geq 1$, $z_k = \boldsymbol{\omega} + \mathbf{A}z_{k-1}$. From (4.16), this so-defined sequence admits a unique limiting point $\lim_{n \rightarrow \infty} z_n = z_{\infty}$. Thus, there exists some $n \in \mathbb{Z}_+^*$ such that for all $k \geq n$, $z_k \in \mathcal{N}_{\infty}$. For such n , we have

$$\begin{aligned} R^n(z; \mathcal{N}_{\infty}) &= \bar{\mathbb{P}}_z(Z_n \in \mathcal{N}_{\infty}) \geq \bar{\mathbb{P}}_z(Z_k = z_k \text{ for all } k = 1, \dots, n) \\ &= \bar{\mathbb{P}}_z(U_0 = \dots = U_{n-1} = 0) > 0. \end{aligned}$$

Next we prove (AG-6). We have for all $n \geq 1$, $u_{0:n-1} \in \mathbf{U}^n$ and $z \in \mathbf{Z}$,

$$(4.17) \quad \Psi \langle u_{0:n-1} \rangle (z) = \mathbf{A}^n z + \sum_{j=0}^{n-1} \mathbf{A}^j (\boldsymbol{\omega} + u_{n-1-j} \mathbf{b}).$$

Then for all $(z, z') \in \mathbf{Z}^2$, $u_{0:p-1} \in \mathbf{U}^p$,

$$\begin{aligned} d_{\mathbf{Z}}(\Psi_{u_0}(z), \Psi_{u_0}(z')) &= |\mathbf{A}(z - z')|_{\infty} \\ &\leq \|\mathbf{A}\|_{\infty} |z - z'|_{\infty} \\ &\leq D d_{\mathbf{Z}}(z, z'), \end{aligned}$$

and

$$\begin{aligned} d_{\mathbf{Z}}(\Psi \langle u_{0:p-1} \rangle (z), \Psi \langle u_{0:p-1} \rangle (z')) &= |\mathbf{A}^p(z - z')|_{\infty} \\ &\leq \|\mathbf{A}^p\|_{\infty} |z - z'|_{\infty} \\ &\leq \rho d_{\mathbf{Z}}(z, z'), \end{aligned}$$

where $D = \|\mathbf{A}\|_{\infty} > 0$ and $\rho = \|\mathbf{A}^p\|_{\infty} < 1$ from (4.8), showing (AG-6). To show (AG-7), we rely on Lemma 10. Let us set $\mathbf{C} = \mathbb{R} = \mathbf{X}$ and $\mathbf{S} = \{1\}$, then the probability measure μ boils down the Dirac mass on \mathbf{S} . For all $(x, u) \in \mathbf{X} \times \mathbf{U}$, let $j(x) = e^{-e^x}$ and $h(x; u) = \frac{e^{x(e^u-1)}}{(e^u-1)!}$. Indeed, h satisfies (H'-1). Hence, from Lemma 10, the instrumental functions α and ϕ are obtained by setting: for all $z, z' \in \mathbf{Z}$,

$$\alpha(z, z') = \frac{e^{-e^{\Pi_p(z) \vee \Pi_p(z')}}}{e^{-e^{\Pi_p(z) \wedge \Pi_p(z')}}} = e^{-|e^{\Pi_p(z)} - e^{\Pi_p(z')}|}$$

and $\phi(z, z') = \Pi_p(z) \wedge \Pi_p(z')$. These so-defined functions α and ϕ indeed satisfy (AG-7)(i). For $(z, z') \in \mathbf{Z}^2$, we have

$$\begin{aligned} 1 - \alpha(z, z') &= 1 - e^{-|e^{\Pi_p(z)} - e^{\Pi_p(z')}|} \leq |e^{\Pi_p(z)} - e^{\Pi_p(z')}| \\ &\leq e^{|\Pi_p(z) \vee \Pi_p(z')|} |\Pi_p(z) - \Pi_p(z')| \\ &\leq W(z, z') \, d_{\mathbf{Z}}(z, z'), \end{aligned}$$

where W is defined by, for all $z, z' \in \mathbf{Z}$,

$$W(z, z') := \gamma_{p-1} (R^{p-1}V_1(z) + R^{p-1}V_1(z'))$$

and $V_1(z) = e^{|z|^\infty}$. Obviously, this so-defined function W satisfies (AG-7)(ii) and (AG-7)(iii). To complete the proof of Proposition 12–(i), it remains to show (AG-7)(iv). Note that (AG-4) holds with $V(z) = e^{\tau|z|^\infty}$ for any $\tau > 0$, thus it also holds with V_1 . For all $z, z' \in \mathbf{Z}$,

$$\begin{aligned} J_1((z, z'); W) &= \int W(\Psi_u(z), \Psi_u(z')) G(\phi(z, z'); du) \\ &= \gamma_{p-1} \int R^{p-1}V_1(\Psi_u(z)) G(\phi(z, z'); du) + \gamma_{p-1} \int V_1(\Psi_u(z')) G(\phi(z, z'); du). \end{aligned}$$

From (4.15), with $m = p - 1$, we have

$$\begin{aligned} (4.18) \quad &\int R^{p-1}V_1(\Psi_u(z)) G(\phi(z, z'); du) \\ &\leq \sum_{\sigma \in \{-1, 1\}} \sum_{k=1}^{2p-1} \sum_{(i_1, \dots, i_{p-1}) \in \{1, 2\}^{p-1}} \beta_{i_1, \dots, i_{p-1}} \times \\ &\quad \int e^{\sigma \delta_k^T (\prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell}) (\omega + \mathbf{A}z + u\mathbf{b})} G(\phi(z, z'); du). \end{aligned}$$

Note that by (4.13), there exist constants c'_1 and c'_2 such that

$$\begin{aligned} &\int e^{\sigma \delta_k^T (\prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell}) \mathbf{b}u} G(\phi(z, z'); du) \\ &\leq c'_1 + c'_2 e^{\sigma \delta_k^T (\prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell}) \mathbf{b}(\delta_p^T z \wedge \delta_p^T z')} \mathbb{1} \left\{ \sigma \delta_k^T \left(\prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell} \right) \mathbf{b} > 0 \right\} \\ &\leq c'_1 + c'_2 e^{\sigma \delta_k^T (\prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell}) \mathbf{b} \delta_p^T z}, \end{aligned}$$

where we have used $\delta_p^T z \wedge \delta_p^T z' \leq \delta_p^T z$ and $\mathbb{1} \left\{ \sigma \delta_k^T \left(\prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell} \right) \mathbf{b} > 0 \right\} \leq 1$. Plugging this inequality into (4.18) and noting that $\mathbf{A}_1 = \mathbf{A}$ and $\mathbf{A}_2 =$

$\mathbf{A} + \mathbf{b}\delta_p^T$, then there exist positive constants γ_{i_1, \dots, i_p} , $(i_1, \dots, i_p) \in \{1, 2\}^p$, such that

$$\begin{aligned} & \int R^{p-1} V_1(\Psi_u(z)) G(\phi(z, z'); du) \\ & \leq \sum_{\sigma \in \{-1, 1\}} \sum_{k=1}^{2p-1} \sum_{(i_1, \dots, i_p) \in \{1, 2\}^p} \gamma_{i_1, \dots, i_p} e^{\sigma \delta_k^T (\prod_{\ell=1}^p \mathbf{A}_{i_\ell}) z} \\ & \leq \left(2(2p-1) \sum_{(i_1, \dots, i_p) \in \{1, 2\}^p} \gamma_{i_1, \dots, i_p} \right) e^{\eta_p |z|_\infty}, \end{aligned}$$

where $\eta_p = \bigvee_{(i_1, \dots, i_p) \in \{1, 2\}^p} \left\| \prod_{\ell=1}^p \mathbf{A}_{i_\ell} \right\|_\infty < 1$. Similarly, we have

$$\int R^{p-1} V_1(\Psi_u(z')) G(\phi(z, z'); du) \leq \left(2(2p-1) \sum_{i_1, \dots, i_p \in \{1, 2\}} \gamma'_{i_1, \dots, i_p} \right) e^{\eta_p |z'|_\infty}$$

for some positive constants $\gamma'_{i_1, \dots, i_p}$, $(i_1, \dots, i_p) \in \{1, 2\}^p$. Thus, there exists $M > 0$ such that for all $z, z' \in \mathbf{Z}$,

$$J_1((z, z'); W) = M \left(e^{\eta_p |z|_\infty} + e^{\eta_p |z'|_\infty} \right).$$

Assumption **(AG-7)(iv)** follows by observing that

$$\lim_{|z|_\infty \vee |z'|_\infty \rightarrow \infty} \frac{J_1((z, z'); W)}{W(z, z')} = 0.$$

This completes the proof of **Proposition 12–(i)**.

Proof of Proposition 12–(ii). As a consequence, we have **(AG-1)**. Assumption **(AG-2)** directly follows. Thus, to show the first part of **Proposition 12–(ii)**, we only need to show that Assumptions **(BG-1)**, **(BG-2)** and **(BG-3)** are satisfied. The second part then follows if we can moreover show that **(a)** and **(b)** in **Theorem 4** hold.

Clearly, **(BG-1)** and **(BG-2)** hold by the definitions of ψ^θ and g^θ given by **(4.4)** and **(4.6)**, respectively. It remains to show **(BG-3)**. Since Θ is compact, then

(4.19)

$$|\omega| \leq \bar{\omega}, |a_i| \leq \bar{a}, |b_i| \leq \bar{b}, \quad \bigvee_{(i_1, \dots, i_p) \in \{1, 2\}^p} \left\| \prod_{\ell=1}^p \mathbf{A}_{i_\ell} \right\|_\infty \leq \bar{\rho}, \|\mathbf{A}\|_\infty \leq L,$$

for some $(\bar{\omega}, \bar{a}, \bar{b}, \bar{\rho}) \in (0, \infty)^3 \times (0, 1)$ and $L > 0$. We set $\mathbf{Z}_1 = \mathbf{Z}$. And so (BG-3)(i) holds. Moreover, for all $(\theta, z, u) \in \Theta \times \mathbf{Z}_1 \times \mathbf{U}$, $g^\theta(\Pi_p(z); y) \leq 1$. Thus, Condition (BG-3)(ii) holds. Now let $z_1 \in \mathbf{Z}$. Using (4.16), (4.17) and (4.19), we get that there exist constants $\bar{c} > 0$ and $\bar{\delta} \in (0, 1)$ such that for all $z \in \mathbf{Z}$, $u_{1:n} \in \mathbf{U}^n$ and $\theta \in \Theta$,

$$\begin{aligned} \left| \Psi^\theta \langle u_{1:n} \rangle(z_1) - \Psi^\theta \langle u_{1:n} \rangle(z) \right|_\infty &= |\mathbf{A}^n(z_1 - z)|_\infty \\ &\leq \bar{c} \bar{\delta}^n |z_1 - z|_\infty. \end{aligned}$$

Thus by (4.12), we get (BG-3)(iii) with

$$\bar{\Psi}(z) = c |z_1 - z|_\infty.$$

Hence (BG-3)(iv) holds and since

$$\left| z_1 - \Psi_u^\theta(z_1) \right| \leq (L+1) |z_1|_\infty + \bar{\omega} + (1 \vee \bar{b})u,$$

we also get (BG-3)(v) provided that

$$(4.20) \quad \bar{\phi}(u) \geq (L+1) |z_1|_\infty + \bar{\omega} + (1 \vee \bar{b})u.$$

It is straightforward to show that for all $\theta \in \Theta$, $(x, x') \in \Pi_p(\mathbf{Z}_1) \times \Pi_p(\mathbf{Z}_1)$ and $u \in \mathbf{U}$,

$$\begin{aligned} \left| \ln g^\theta(x; u) - \ln g^\theta(x'; u) \right| &\leq |x - x'| e^{|x| \vee |x'|} e^u \\ &\leq |x - x'| e^{|x - \Pi_p(z_1)| \vee |x' - \Pi_p(z_1)|} e^{u + |\Pi_p(z_1)|}. \end{aligned}$$

Thus for all $\theta \in \Theta$, $(z, z') \in \mathbf{Z}_1 \times \mathbf{Z}_1$ and $u \in \mathbf{U}$,

$$\left| \ln g^\theta(x; u) - \ln g^\theta(x'; u) \right| \leq |z - z'|_\infty e^{|z - z_1|_\infty \vee |z' - z_1|_\infty} e^{u + |\Pi_p(z_1)|}.$$

We thus obtain (BG-3)(v), (BG-3)(vi) and (BG-3)(vii) by setting $C = 1$,

$$H(s) = s, \quad s \in \mathbb{R}_+,$$

and

$$\bar{\phi}(u) = (L+1) |z_1|_\infty + \bar{\omega} + (1 \vee \bar{b})u + e^{u + |\Pi_p(z_1)|}.$$

In addition, for all $\theta \in \Theta$, $x \in \mathbf{X}$ and $\tau > 0$, we have

$$\int e^{\tau u} G^\theta(x; du) \leq 1 + 2^\tau e^{\tau x}.$$

Hence, by letting $\bar{V} = V$ and using that for all $\theta \in \Theta$, $\pi_1^\theta V < \infty$, with the above definitions, we obtain (BG-3)(viii). This establishes the equivalence-class consistency of the MLE for the log-linear Poisson GARCH(p, p).

To conclude the strong consistency of the MLE, it remains to check (a) and (b) in Theorem 4. Condition (a) clearly holds by the identifiability of the Poisson distribution. Note that as a consequence of the earlier proof, we have, for all $\theta \in \Theta$, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\Psi^\theta \langle U_{-\infty:0} \rangle = \sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + U_{-k} \mathbf{b}).$$

By stationarity, for all $\theta \in \Theta$ and $t \in \mathbb{Z}$, we have, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\Psi^\theta \langle U_{-\infty:t} \rangle = \sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + U_{-k+t} \mathbf{b}).$$

Moreover, if

$$\Psi^\theta \langle U_{-\infty:0} \rangle = \Psi^{\theta_\star} \langle U_{-\infty:0} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

then by stationarity, for all $t \in \mathbb{Z}$,

$$\Psi^\theta \langle U_{-\infty:t} \rangle = \Psi^{\theta_\star} \langle U_{-\infty:t} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

and thus

$$(4.21) \quad \sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + U_{-k+t} \mathbf{b}) = \sum_{k=0}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + U_{-k+t} \mathbf{b}_\star), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

This implies

$$(4.22) \quad (\mathbf{b} - \mathbf{b}_\star) U_t = \sum_{k=1}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + U_{-k+t} \mathbf{b}_\star) - \sum_{k=1}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + U_{-k+t} \mathbf{b}), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Conditionally on $\sigma(U_{-\infty:t-1})$, $(e^{U_t} - 1)$ is a Poisson variable with positive intensity. Thus the left-hand side of (4.22) is $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. constant only if $\mathbf{b} = \mathbf{b}_\star$, implying $b_1^\star = b_1$. Then, (4.21) reduces to

$$(4.23) \quad \sum_{k=1}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + U_{-k+t} \mathbf{b}) = \sum_{k=1}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + U_{-k+t} \mathbf{b}_\star), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Thus by repeating the same argument as to obtain $\mathbf{b} = \mathbf{b}_\star$, (4.23) and so on yield that, for all integer $k \geq 1$,

$$(\mathbf{A}^k - \mathbf{A}_\star^k) \mathbf{b}_\star = 0.$$

And by (4.10), we have $\mathbf{A} = \mathbf{A}_*$, yielding that $(a_{1:p}, b_{2:p}) = (a_{1:p}^*, b_{2:p}^*)$. It is immediate from (4.21) that $\boldsymbol{\omega} = \boldsymbol{\omega}_*$, and this is equivalent to having that $\omega = \omega^*$. The strong consistency therefore follows.

Proof of Proposition 12–(iii). By Theorem 7, it suffices to show that (CG-1)–(CG-7) hold. From the definition of $\Psi_u^\theta(z)$, we have $a(\theta, u) = \mathbf{A}$ and $b(\theta, u) = \boldsymbol{\omega} + u\mathbf{b}$, both admitting linear forms in θ . Thus from compactness of Θ and (4.8) and by noting that for all $\tau > 0$,

$$(4.24) \quad \tilde{\mathbb{E}}^{\theta_*} [e^{\tau U_1}] < \infty,$$

it follows that (CG-1) holds. Assumptions (CG-2) and (CG-3) clearly hold. Assumption (CG-4) is immediate as the conclusion of Proposition 12–(ii) above. We have

$$\begin{aligned} \ell_{z,k}^\theta &= \ln g^\theta \left(\Pi_p(\Psi^\theta \langle U_{1:k-1} \rangle(z)); U_k \right) \\ &= (e^{U_k} - 1) \boldsymbol{\delta}_p^T \Psi^\theta \langle U_{1:k-1} \rangle(z) - e^{\boldsymbol{\delta}_p^T \Psi^\theta \langle U_{1:k-1} \rangle(z)} - \ln(e^{U_k} - 1)!, \end{aligned}$$

$$\begin{aligned} \ell_k^\theta &= \ln g^\theta \left(\Pi_p(\Psi^\theta \langle U_{-\infty:k-1} \rangle(z)); U_k \right) \\ &= (e^{U_k} - 1) \boldsymbol{\delta}_p^T \Psi^\theta \langle U_{-\infty:k-1} \rangle - e^{\boldsymbol{\delta}_p^T \Psi^\theta \langle U_{-\infty:k-1} \rangle} - \ln(e^{U_k} - 1)!. \end{aligned}$$

Then from Lemma 5, for all i , $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.,

$$\frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} = \left(e^{U_k} - 1 - e^{\boldsymbol{\delta}_p^T \Psi^\theta \langle U_{1:k-1} \rangle(z)} \right) \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{1:k-1} \rangle(z)$$

$$\frac{\partial \ell_k^\theta}{\partial \theta_i} = \left(e^{U_k} - 1 - e^{\boldsymbol{\delta}_p^T \Psi^\theta \langle U_{-\infty:k-1} \rangle} \right) \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:k-1} \rangle$$

Thus for all i , $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.,

$$\begin{aligned} \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| &\leq \left| \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:k-1} \rangle - \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{1:k-1} \rangle(z) \right| (e^{U_k} - 1) \\ &+ \left| e^{\boldsymbol{\delta}_p^T \Psi^\theta \langle U_{-\infty:k-1} \rangle} - e^{\boldsymbol{\delta}_p^T \Psi^\theta \langle U_{1:k-1} \rangle(z)} \right| \left| \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:k-1} \rangle \right| \\ &+ e^{\boldsymbol{\delta}_p^T \Psi^\theta \langle U_{1:k-1} \rangle(z)} \left| \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:k-1} \rangle - \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{1:k-1} \rangle(z) \right| \end{aligned}$$

Note that for all $z \in \mathbf{Z}$ and $k \geq 1$, we have, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\Psi^\theta \langle U_{-\infty:k-1} \rangle - \Psi^\theta \langle U_{1:k-1} \rangle(z) = \mathbf{A}^{k-1} \left(\Psi^\theta \langle U_{-\infty:0} \rangle - z \right),$$

and for all i ,

$$\begin{aligned} \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:k-1} \rangle - \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{1:k-1} \rangle(z) \\ = \frac{\partial(\mathbf{A}^{k-1})}{\partial \theta_i} \left(\Psi^\theta \langle U_{-\infty:0} \rangle - z \right) + \mathbf{A}^{k-1} \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:0} \rangle. \end{aligned}$$

By similar arguments as in the proof of [Lemma 5](#) (see [Section 6.1](#)) and noting that $\sup_{\theta \in \Theta} |\Psi^\theta \langle U_{-\infty:0} \rangle|_\infty$ and $\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta} \Psi^\theta \langle U_{-\infty:0} \rangle \right|_\infty$ are finite $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s., there exist a constant $\rho_1 \in (0, 1)$ and a $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. finite random variable $\tilde{M}_1 = \tilde{M}_1(z)$ such that for all $k \geq 1$, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\left| \Psi^\theta \langle U_{-\infty:k-1} \rangle - \Psi^\theta \langle U_{1:k-1} \rangle(z) \right|_\infty \leq \tilde{M}_1 \rho_1^k,$$

and for all i , $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\left| \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:k-1} \rangle - \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{1:k-1} \rangle(z) \right|_\infty \leq \tilde{M}_1 \rho_1^k.$$

From [Lemma 5](#), we also have that, for all i ,

$$(4.25) \quad \tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:0} \rangle \right|_\infty \right] < \infty.$$

Recall that for all $t, t' \in \mathbb{R}$, $|e^t - e^{t'}| \leq e^{|t \vee t'|} |t - t'|$. Thus, we deduce that

$$\begin{aligned} \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| \\ \leq e^{\tilde{M}_1} \rho_1^k \tilde{M}_1 \left[e^{U_k} - 1 + e^{|\Psi^\theta \langle U_{-\infty:k-1} \rangle|_\infty} \left(1 + \left| \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:k-1} \rangle \right|_\infty \right) \right]. \end{aligned}$$

On the other hand, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s., we have

$$e^{|\Psi^\theta \langle U_{-\infty:0} \rangle|_\infty} \leq e^{\sup_{\theta \in \Theta} |\Psi^\theta \langle U_{-\infty:0} \rangle|_\infty} \leq e^{\sum_{k=0}^\infty \sup_{\theta \in \Theta} |\mathbf{A}^k(\boldsymbol{\omega} + U_{-k} \mathbf{b})|_\infty}.$$

From [\(4.16\)](#) and [\(4.19\)](#), there exist constants $c_1, c_2 > 0$ and $\bar{\delta} \in (0, 1)$ such that, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$e^{\sup_{\theta \in \Theta} |\Psi^\theta \langle U_{-\infty:0} \rangle|_\infty} \leq c_1 e^{c_2 \sum_{k=0}^\infty \bar{\delta}^k U_{-k}} \leq c_1 (1 - \bar{\delta}) \sum_{k=0}^\infty \bar{\delta}^k e^{c_2 (1 - \bar{\delta})^{-1} U_{-k}},$$

and from (4.24), it follows that for any positive $t > 0$,

$$(4.26) \quad \tilde{\mathbb{E}}^{\theta^*} \left[e^{t \sup_{\theta \in \Theta} |\Psi^\theta \langle U_{-\infty:0} \rangle|_\infty} \right] < \infty.$$

Thus, by stationarity and from (4.25) and (4.26), Lemma 21 assures that there exist a constant $\rho_2 \in (0, 1)$ and a $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. finite random variable $\tilde{M}_2 = \tilde{M}_2(z)$ such that for all i and for all $k \geq 1$, $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| \leq \rho_2^k \tilde{M}_2.$$

Therefore, Assumption (CG-5) holds. We now check (CG-6). We have

$$\begin{aligned} \ell_1^\theta &= \ln g^\theta \left(\Pi_p(\Psi^\theta \langle U_{-\infty:0} \rangle); U_1 \right) \\ &= (e^{U_1} - 1) \delta_p^T \Psi^\theta \langle U_{-\infty:0} \rangle - e^{\delta_p^T \Psi^\theta \langle U_{-\infty:0} \rangle} - \ln(e^{U_1} - 1)!. \end{aligned}$$

Then for all i and j , we have

$$\frac{\partial \ell_1^\theta}{\partial \theta_i} = \left(e^{U_1} - 1 - e^{\delta_p^T \Psi^\theta \langle U_{-\infty:0} \rangle} \right) \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:0} \rangle$$

and from Lemma 5, $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned} \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_j} &= \left(e^{U_1} - 1 - e^{\delta_p^T \Psi^\theta \langle U_{-\infty:0} \rangle} \right) \delta_p^T \frac{\partial^2}{\partial \theta_i \partial \theta_j} \Psi^\theta \langle U_{-\infty:0} \rangle \\ &\quad + e^{\delta_p^T \Psi^\theta \langle U_{-\infty:0} \rangle} \left(\delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:0} \rangle \right) \left(\delta_p^T \frac{\partial}{\partial \theta_j} \Psi^\theta \langle U_{-\infty:0} \rangle \right). \end{aligned}$$

Using similar arguments as the proof of Lemma 5 together with (4.24), we can show that for all i and j , we have

$$\tilde{\mathbb{E}}^{\theta^*} \left[\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:0} \rangle \right|_\infty^2 \right] < \infty$$

and

$$\tilde{\mathbb{E}}^{\theta^*} \left[\sup_{\theta \in \Theta} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \Psi^\theta \langle U_{-\infty:0} \rangle \right|_\infty \right] < \infty.$$

By (4.26) and (4.24), we can directly show that, for all i and j ,

$$\tilde{\mathbb{E}}^{\theta^*} \left[\sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right|^2 \right] < \infty$$

and

$$\tilde{\mathbb{E}}^{\theta^*} \left[\sup_{\theta \in \Theta} \left| \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_j} \right| \right] < \infty.$$

Thus, (CG-6)-(i) and (CG-6)-(ii) hold. Note that for all i, j , we have, $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\frac{\partial}{\partial \theta_i} g^\theta \left(\Pi_p(\Psi^\theta \langle U_{-\infty:0} \rangle); U_1 \right) = g^\theta \left(\Pi_p(\Psi^\theta \langle U_{-\infty:0} \rangle); U_1 \right) \frac{\partial \ell_1^\theta}{\partial \theta_i}$$

and

$$\begin{aligned} & \frac{\partial^2}{\partial \theta_i \partial \theta_j} g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); U_1 \right) \\ &= g^\theta \left(\Pi_p(\Psi^\theta \langle U_{-\infty:0} \rangle); U_1 \right) \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_j} + \frac{\partial}{\partial \theta_i} g^\theta \left(\Pi_p(\Psi^\theta \langle U_{-\infty:0} \rangle); Y_1 \right) \frac{\partial \ell_1^\theta}{\partial \theta_j} \end{aligned}$$

Then, (CG-6)-(iii) follows by observing that for all i , $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \frac{\partial}{\partial \theta_i} g^\theta \left(\Pi_p(\Psi^\theta \langle U_{-\infty:0} \rangle); U_1 \right) \right| = g^\theta \left(\Pi_p(\Psi^\theta \langle U_{-\infty:0} \rangle); U_1 \right) \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right| \leq \sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right|$$

and

$$\tilde{\mathbb{E}}^{\theta^*} \left[\sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right| \right] \leq \left(\tilde{\mathbb{E}}^{\theta^*} \left[\sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right|^2 \right] \right)^{1/2} < \infty.$$

Similarly, (CG-6)-(iv) holds. To complete the proof of asymptotic normality, it remains to check (CG-7). By Remark 4, it is sufficient to show that (3.23) holds. We establish this fact by following the argument provided by Francq and Zakoian (2004). Now let $\alpha = (\alpha_0, \dots, \alpha_{2p}) \in \mathbb{R}^{2p+1}$ and for convenience, write $\theta = (\theta_0, \dots, \theta_{2p}) = (\omega, a_{1:p}, b_{1:p})$. We have, $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\alpha^T \frac{\partial \ell_1^{\theta^*}}{\partial \theta} = \sum_{i=0}^{2p} \alpha_i \left(e^{U_1} - 1 - e^{\delta_p^T \Psi^{\theta^*} \langle U_{-\infty:0} \rangle} \right) \delta_p^T \frac{\partial \Psi^{\theta^*} \langle U_{-\infty:0} \rangle}{\partial \theta_i}.$$

Since $(e^{U_1} - 1)$ conditionally on $\sigma(U_{-\infty:0})$ is a Poisson variable with positive intensity, then $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., $\alpha^T \frac{\partial \ell_1^{\theta^*}}{\partial \theta} = 0$ implies

$$\sum_{i=0}^{2p} \alpha_i \delta_p^T \frac{\partial \Psi^{\theta^*} \langle U_{-\infty:0} \rangle}{\partial \theta_i} = 0, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

By stationarity, we have for all $t \in \mathbb{Z}$,

$$(4.27) \quad \sum_{i=0}^{2p} \alpha_i \boldsymbol{\delta}_p^T \frac{\partial \Psi^{\theta_*} \langle U_{-\infty:t} \rangle}{\partial \theta_i} = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

In view of (4.3) and by noting that $\tilde{\mathbb{P}}^{\theta_*}$ -a.s., $X_t^\theta = \boldsymbol{\delta}_p^T \Psi^\theta \langle U_{-\infty:t} \rangle$ for all $\theta \in \Theta$ and $t \in \mathbb{Z}$, we have from (4.27) that for all $t \in \mathbb{Z}$,

$$0 = \boldsymbol{\alpha}^T \frac{\partial X_t^{\theta_*}}{\partial \theta} = \boldsymbol{\alpha}^T \begin{pmatrix} 1 \\ X_{t-1}^{\theta_*} \\ \vdots \\ X_{t-p}^{\theta_*} \\ U_{t-1} \\ \vdots \\ U_{t-p} \end{pmatrix} + \sum_{i=1}^p a_i \boldsymbol{\alpha}^T \frac{\partial X_{t-i}^{\theta_*}}{\partial \theta} = \boldsymbol{\alpha}^T \begin{pmatrix} 1 \\ X_{t-1}^{\theta_*} \\ \vdots \\ X_{t-p}^{\theta_*} \\ U_{t-1} \\ \vdots \\ U_{t-p} \end{pmatrix}, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

Since $(e^{U_{t-1}} - 1)$ conditionally on $\sigma(U_{-\infty:t-2})$ is a Poisson variable with positive intensity, we then have that $\alpha_{p+1} = 0$. By the same reason, it can be shown that $\alpha_{p+2:p+2+i} = 0$ if $\alpha_{2:2+i} = 0$, for $i \in \{0, \dots, p-2\}$. Thus, if $\boldsymbol{\alpha} \neq 0$ entails an ODM($p-1, p-1$) representation, which is impossible since the model is identifiable. Hence, $\boldsymbol{\alpha} = 0$ and this completes the proof. \square

Our next example is the NBIN-GARCH(p, p), $p \in \mathbb{Z}_+^*$, which is defined as follows.

4.2. NBIN-GARCH(p, p) Model.

EXAMPLE 2 (Definition of NBIN-GARCH(p, p)). Consider a statistical model recursively defined by: for all $k \in \mathbb{Z}_+$,

$$(4.28) \quad \begin{aligned} X_{k+1} &= \omega + \sum_{i=1}^p a_i X_{k-i+1} + \sum_{i=1}^p b_i Y_{k-i+1}, \\ Y_{k+1} \mid \mathcal{F}_k &\sim \mathcal{NB} \left(r, \frac{X_{k+1}}{1 + X_{k+1}} \right), \end{aligned}$$

where

– the parameter space

$$\Theta \subset \{ \theta = (\omega, a_{1:p}, b_{1:p}, r) \in \mathbb{R}_+^* \times \mathbb{R}_+^p \times \mathbb{R}_+^p \times \mathbb{R}_+^* \},$$

- $\mathcal{F}_k = \sigma(X_{-p+1:k+1}, Y_{-p+1:k})$,
- $\mathbf{X} = (0, \infty)$, $\mathcal{X} = \mathcal{B}(\mathbf{X})$, $\mathbf{Y} = \mathbb{Z}_+$, $\mathcal{Y} = \mathcal{B}(\mathbf{Y})$ and the metrics on \mathbf{X} and \mathbf{Y} are any metrics on \mathbb{R} and \mathbb{Z} , respectively.

Denoting for all $(x_{1:p}, y_{1:p}) \in \mathbf{X}^p \times \mathbf{Y}^p$,

$$(4.29) \quad \psi_{y_{1:p}}^\theta(x_{1:p}) := \omega + \sum_{i=1}^p a_i x_{p-i+1} + \sum_{i=1}^p b_i y_{p-i+1},$$

Model (4.28) can be expressed as the standard form of Model (4.30) below:

$$(4.30) \quad \begin{aligned} X_{k+1} &= \psi_{Y_{k-p+1:k}}^\theta(X_{k-p+1:k}), \\ Y_{k+1} | \mathcal{F}_k &\sim \mathcal{NB}\left(r, \frac{X_{k+1}}{1 + X_{k+1}}\right). \end{aligned}$$

The density g^θ and the probability law G^θ of Y_k conditional on $X_k = x$ are given by, for all $y \in \mathbf{Y}$ and $A \in \mathcal{Y}$,

$$(4.31) \quad g^\theta(x; y) = \frac{\Gamma(r+y)}{y! \Gamma(r)} \left(\frac{1}{1+x}\right)^r \left(\frac{x}{1+x}\right)^y,$$

and

$$(4.32) \quad G^\theta(x; A) = \sum_{y \in A} \mathbb{1}_A(y) g^\theta(x; y).$$

According to Definition 1, Model (4.28) is an ODM(p, p) and is dominated by the counting measure on \mathbb{Z} . Now, let $\mathbf{Z} = \mathbf{X}^p \times \mathbf{Y}^{p-1}$ and $\mathcal{Z} = \mathcal{X}^{\otimes p} \otimes \mathcal{Y}^{\otimes(p-1)}$ and define the measurable function Ψ^θ similarly as in (2.6), which is given by, for all $(z, y) \in \mathbf{Z} \times \mathbf{Y}$,

$$(4.33) \quad \Psi_y^\theta(z) = \omega + \mathbf{A}z + y\mathbf{b}.$$

The process $\{Z_k : k \in \mathbb{Z}_+\}$, where $Z_k = (X_{k-p+1:k}, Y_{k-p+1:k-1})$, is a Markov chain on $(\mathbf{Z}, \mathcal{Z})$ with transition kernel R^θ given by (2.11) with Ψ^θ and G^θ defined in (4.32) and (4.33), respectively. Let us now state our ergodicity, consistency and asymptotic normality results for the NBIN-GARCH(p, p) model.

PROPOSITION 14. *Suppose that $\theta \in \Theta$ satisfies*

$$(4.34) \quad |\lambda|_{\max}(\mathbf{A} + r\mathbf{b}\delta_p^T) < 1.$$

- (i) Then the conclusion of [Theorem 9](#) holds with $\bar{V}(z) = \mathbf{1}^T z$ for all $z \in \mathbf{Z}$.
(ii) [Theorem 3](#) holds with any $z_1 \in \mathbf{Z}$. If the true parameter $\theta_\star = (\omega^\star, a_{1:p}^\star, b_{1:p}^\star)$ moreover satisfies: for all $k \in \mathbb{Z}_+^*$,

$$(4.35) \quad (\mathbf{A}^k - \mathbf{A}_\star^k) \mathbf{b}_\star = 0 \quad \text{implies} \quad \mathbf{A} = \mathbf{A}_\star,$$

then [Theorem 4](#) also holds.

REMARK 11. From [Lemma 11](#), Condition (4.34) is equivalent to $\sum_{i=1}^p (a_i + r b_i) < 1$.

PROOF. **Proof of Proposition 14–(i).** In this part, for convenience, let us drop the super script θ from for example G^θ , R^θ , etc. Now set $V = \bar{V}$. From [Theorem 9](#), we need to show that (AG-3), (AG-4), (AG-5), (AG-6) and (AG-7) hold. Assumption (AG-3) holds with any metric dz associated to a norm on the finite dimensional space \mathbf{Z} so that \mathcal{Z} is the associated Borel σ -field (The precise choice of this metric is postponed to the verification of (AG-6)). The fact that R is weak Feller follows by observing that for fixed $r > 0$, as $p \rightarrow p'$, $\mathcal{NB}(r, p)$ converges weakly to $\mathcal{NB}(r, p')$. Moreover, we have, for all $z \in \mathbf{Z}$,

$$\begin{aligned} RV(z) &= \int V(\omega + \mathbf{A}z + y\mathbf{b}) G(\Pi_p(z); dy) \\ &= \mathbf{1}^T \omega + \mathbf{1}^T (\mathbf{A} + r\mathbf{b}\delta_p^T) z. \end{aligned}$$

By iteration, we obtain that for all $k \in \mathbb{Z}_+^*$,

$$R^k V(z) = \mathbf{1}^T \omega + \dots + \mathbf{1}^T (\mathbf{A} + r\mathbf{b}\delta_p^T)^{k-1} \omega + \mathbf{1}^T (\mathbf{A} + r\mathbf{b}\delta_p^T)^k z.$$

Since $|\lambda|_{\max}(\mathbf{A} + r\mathbf{b}\delta_p^T) < 1$, then there exists a pair $(q, \rho) \in \mathbb{Z}_+^* \times (0, 1)$ such that

$$R^q V(z) \leq \mathbf{1}^T \omega + \dots + \mathbf{1}^T (\mathbf{A} + r\mathbf{b}\delta_p^T)^{q-1} \omega + \rho^q \mathbf{1}^T z.$$

Hence (AG-4) holds with some $q \geq 1$, $\lambda = \rho^q < 1$ and $\beta = \mathbf{1}^T \omega + \dots + \mathbf{1}^T (\mathbf{A} + r\mathbf{b}\delta_p^T)^{q-1} \omega \geq 0$. We now show (AG-5). Let us use the norm

$$\|\mathbf{M}\| = \bigvee_{j=1}^{2p-1} \sum_{i=1}^{2p-1} |\mathbf{M}_{i,j}| = \sup_{|z| \leq 1} |\mathbf{M}z|$$

on $(2p-1) \times (2p-1)$ matrices, where $|z| = \sum_{i=1}^{2p-1} |z_i|$ is a norm of $z = z_{1:2p-1} \in \mathbf{Z}$. Note that from (4.34), there exists $\delta \in (0, 1)$ and $c > 0$ such that, for any $k \geq 1$,

$$(4.36) \quad \left\| \mathbf{A}^k \right\| \leq c \delta^k .$$

Hence $(\mathbf{I}_{2p-1} - \mathbf{A})^{-1} = \mathbf{I}_{2p-1} + \sum_{k \geq 1} \mathbf{A}^k$ is well defined and we set $z_\infty = (\mathbf{I}_{2p-1} - \mathbf{A})^{-1} \boldsymbol{\omega}$. Let \mathcal{N}_∞ be an open neighborhood of z_∞ and let $z \in \mathbf{Z}$. Define recursively the sequence $z_0 = z$ and for all $k \geq 1$, $z_k = \boldsymbol{\omega} + \mathbf{A}z_{k-1}$. From (4.36), this so-defined sequence admits a unique limiting point $\lim_{n \rightarrow \infty} z_n = z_\infty$. Thus, there exists some $n \in \mathbb{Z}_+^*$ such that for all $k \geq n$, $z_k \in \mathcal{N}_\infty$. For such n , we have

$$\begin{aligned} R^n(z, \mathcal{N}_\infty) &= \bar{\mathbb{P}}_z(Z_n \in \mathcal{N}_\infty) \geq \bar{\mathbb{P}}_z(Z_k = z_k \text{ for all } k = 1, \dots, n) \\ &= \bar{\mathbb{P}}_z(Y_0 = \dots = Y_{n-1} = 0) > 0. \end{aligned}$$

Next we prove (AG-6). We have

$$\Psi_y(z) - \Psi_y(z') = \mathbf{A}(z - z').$$

Since (4.34) implies that $|\lambda|_{\max}(\mathbf{A}) < 1$, there exists a vector norm $|\cdot|_Z$, which makes \mathbf{A} strictly contracting. Choosing the metric d_Z on \mathbf{Z} (embedded in \mathbb{R}^{2p-1}) as the one derived from this norm, we get (AG-6). To show (AG-7), we rely on Lemma 10. Let us set $\mathbf{C} = (0, \infty) = \mathbf{X}$ and $\mathbf{S} = \{1\}$, then the probability measure μ boils down the Dirac mass on \mathbf{S} . For all $(x, y) \in \mathbf{X} \times \mathbf{Y}$, let $j(x) = (1+x)^{-r}$ and $h(x; y) = \frac{\Gamma(r+y)}{y! \Gamma(r)} \left(\frac{x}{1+x}\right)^y$. Indeed, h satisfies (H-1). Hence, from lemma Lemma 10 we α and ϕ are obtained by setting: for all $z, z' \in \mathbf{Z}$,

$$\alpha(z, z') = \left(\frac{1 + \Pi_p(z) \wedge \Pi_p(z')}{1 + \Pi_p(z) \vee \Pi_p(z')} \right)^r$$

and

$$\phi(z, z') = \Pi_p(z) \wedge \Pi_p(z').$$

These so-defined functions α and ϕ indeed satisfy (AG-7)(i). For any given $r > 0$, for all $(z, z') \in \mathbf{Z}^2$, we have

$$\begin{aligned} 1 - \alpha(z, z') &= 1 - \left(\frac{1 + \Pi_p(z) \wedge \Pi_p(z')}{1 + \Pi_p(z) \vee \Pi_p(z')} \right)^r \\ &\leq (1 \vee r) |\Pi_p(z) - \Pi_p(z')| \\ &\leq W(z, z') d_Z(z, z'), \end{aligned}$$

where d_Z is the metric previously defined and W is defined by $W(z, z') = c_{d_Z}(1 \vee r)$ with $c_{d_Z} > 0$ is conveniently chosen (such a constant exists since d_Z is the metric associated to a norm and \mathbf{Z} is a subset of finite dimensional space). Thus, (AG-7)(ii) and (AG-7)(iii) hold and, since for all $y \in \mathbf{Y}$ and $z \in \mathbf{Z}$, $\Pi_p(\Psi_y(z)) \geq \omega$, $W(\Psi_y(z), \Psi_y(z'))$ is uniformly bounded over $(z, z', y) \in \mathbf{Z} \times \mathbf{Z} \times \mathbf{Y}$ and (AG-7)(iv) holds. This completes

the proof.

Proof of Proposition 14–(ii). Note that (AG-2) immediately follows. Thus, to show the first part of Proposition 12–(ii), we only need to show that Assumptions (BG-1), (BG-2) and (BG-3) are satisfied. The second part then follows if we can moreover show that (a) and (b) in Theorem 4 hold.

Clearly, (BG-1) and (BG-2) hold by the definitions of ψ^θ and g^θ given by (4.29) and (4.31), respectively. It remains to show (BG-3). Since Θ is compact, then

$$(4.37) \quad \underline{\omega} \leq \omega \leq \bar{\omega}, 0 \leq a_i \leq \bar{a}, 0 \leq b_i \leq \bar{b}, |\lambda|_{\max}(\mathbf{A} + r\mathbf{b}\delta_p^T) \leq \bar{\rho}, \underline{r} \leq r \leq \bar{r}$$

for some $(\underline{\omega}, \bar{\omega}, \bar{a}, \bar{b}, \underline{r}, \bar{r}, \bar{\rho}) \in (0, \infty)^6 \times (0, 1)$.

We set $\mathbf{Z}_1 = \mathbf{Z}$ so that (BG-3)(i) holds. Moreover, for all $(\theta, z, y) \in \Theta \times \mathbf{Z}_1 \times \mathbf{Y}$, $g^\theta(\Pi_p(z); y) \leq 1$. Thus, Condition (BG-3)(ii) holds. Note that we have, for all $n \in \mathbb{Z}_+^*$ and $z \in \mathbf{Z}$,

$$(4.38) \quad \Psi^\theta \langle y_{0:n-1} \rangle(z) = \mathbf{A}^n z + \sum_{j=0}^{n-1} \mathbf{A}^j (\boldsymbol{\omega} + y_{n-1-j} \mathbf{b}).$$

Now let $|\cdot|$ be the Taxicab norm on \mathbf{Z} and let $z_1 \in \mathbf{Z}$. Using (4.36), (4.37) and (4.38), we get that there exist constants $\bar{c} > 0$ and $\bar{\delta} \in (0, 1)$ such that for all $z \in \mathbf{Z}$, $y_{1:n} \in \mathbf{Y}^n$ and $\theta \in \Theta$,

$$\begin{aligned} d_{\mathbf{Z}} \left(\Psi^\theta \langle y_{1:n} \rangle(z_1), \Psi^\theta \langle y_{1:n} \rangle(z) \right) &= |\mathbf{A}^n(z_1 - z)|_{\mathbf{Z}} \\ &\leq \bar{c} \bar{\delta}^n |z_1 - z|. \end{aligned}$$

Thus, we get (BG-3)(iii) with

$$\bar{\Psi}(z) = \bar{c} |z_1 - z|.$$

Hence (BG-3)(iv) holds and since

$$\left| z_1 - \Psi_y^\theta(z_1) \right| \leq (L + 1) |z_1| + \bar{\omega} + (1 + \bar{b})y,$$

for some $L > 0$, we also get (BG-3)(v) provided that for all $y \in \mathbf{Y}$,

$$(4.39) \quad \bar{\phi}(y) \geq (L + 1) |z_1| + \bar{\omega} + (1 + \bar{b})y.$$

For all $\theta \in \Theta$, $(x, x') \in \Pi_p(\mathbf{Z}_1) \times \Pi_p(\mathbf{Z}_1)$ and $y \in \mathbf{Y}$,

$$\begin{aligned} \left| \ln g^\theta(x; y) - \ln g^\theta(x'; y) \right| &= |(r + y)[\ln(1 + x') - \ln(1 + x)] + y[\ln x - \ln x']| \\ &\leq [(r + y)(1 + \underline{\omega})^{-1} + y \underline{\omega}^{-1}] |x - x'| \\ &\leq [\bar{r} + y(1 + \underline{\omega}^{-1})] |x - x'|. \end{aligned}$$

We thus obtain (BG-3)(v), (BG-3)(vi) and (BG-3)(vii) by setting $C = 0$, $H(s) = s$ and $\bar{\phi}(y) = \bar{\omega} \vee \bar{r} + (1 + L)|z_1| + (\bar{b} + 1 + \underline{\omega}^{-1})y$. Observe that for all $z \in \mathbb{Z}$,

$$\int \ln^+ y G^\theta(\Pi_p(z); dy) \leq \int y G^\theta(\Pi_p(z); dy) = r\Pi_p(z) \leq \bar{r}\bar{V}(z).$$

Hence, using that for all $\theta \in \Theta$, $\pi_1^\theta V < \infty$, we obtain (BG-3)(viii), establishing the equivalence-class consistency of the MLE for the NBIN-GARCH(p, p) model.

To conclude the strong consistency of the MLE, it remains to check (a) and (b) in Theorem 4. Condition (a) clearly holds by the identifiability of the negative binomial distribution. Note that as a consequence of the earlier proof, we have, for all $\theta \in \Theta$, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\Psi^\theta \langle Y_{-\infty:0} \rangle = \sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + Y_{-k}\mathbf{b}).$$

By stationarity, for all $\theta \in \Theta$ and $t \in \mathbb{Z}$, we have, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\Psi^\theta \langle Y_{-\infty:t} \rangle = \sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + Y_{-k+t}\mathbf{b}).$$

Moreover, if

$$\Psi^\theta \langle Y_{-\infty:0} \rangle = \Psi^{\theta_\star} \langle Y_{-\infty:0} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

then by stationarity, for all $t \in \mathbb{Z}$,

$$\Psi^\theta \langle Y_{-\infty:t} \rangle = \Psi^{\theta_\star} \langle Y_{-\infty:t} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

and thus

$$(4.40) \quad \sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + Y_{-k+t}\mathbf{b}) = \sum_{k=0}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + Y_{-k+t}\mathbf{b}_\star), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

This implies

$$(4.41) \quad (\mathbf{b} - \mathbf{b}_\star)Y_t = \sum_{k=1}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + Y_{-k+t}\mathbf{b}_\star) - \sum_{k=1}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + Y_{-k+t}\mathbf{b}), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Conditionally on $\sigma(Y_{-\infty:t-1})$, Y_t is negative binomial variable with parameter $(r_\star, p_\star) \in (0, \infty) \times (0, 1)$. Thus, the left-hand side of (4.41) is $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. constant only if $\mathbf{b} = \mathbf{b}_\star$, implying $b_1^\star = b_1$. Then, (4.40) reduces to

$$(4.42) \quad \sum_{k=1}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + Y_{-k+t}\mathbf{b}) = \sum_{k=1}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + Y_{-k+t}\mathbf{b}_\star), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Thus, by repeating the same argument as to obtain $\mathbf{b} = \mathbf{b}_*$, (4.42) and so on yield that, for all integer $k \geq 1$,

$$(\mathbf{A}^k - \mathbf{A}_*^k)\mathbf{b}_* = 0.$$

And by (4.35), we have $\mathbf{A} = \mathbf{A}_*$, yielding that $(a_{1:p}, b_{2:p}) = (a_{1:p}^*, b_{2:p}^*)$. It is immediate from (4.40) that $\boldsymbol{\omega} = \boldsymbol{\omega}_*$, and this is equivalent to having that $\omega = \omega^*$. Thus, (b) is satisfied and the strong consistency therefore follows. \square

PROPOSITION 15. *Suppose that $\theta \in \Theta$ satisfies*

$$(4.43) \quad |\lambda|_{\max} (\mathbf{A}^T \mathbf{A} + r (\mathbf{A}^T \mathbf{b} \boldsymbol{\delta}_p^T + \boldsymbol{\delta}_p \mathbf{b}^T \mathbf{A}) + r(r+1) \boldsymbol{\delta}_p \mathbf{b}^T \mathbf{b} \boldsymbol{\delta}_p^T) < 1.$$

In addition, assume that the true parameter θ_ lies in the interior of Θ and satisfies (4.35), then Theorem 7 holds.*

REMARK 12. Condition (4.43) implies Condition (4.34). Thus the stationarity and ergodicity for the model follow. If moreover (4.35) is satisfied, then the strong consistency, that is, $\lim_{n \rightarrow \infty} \hat{\theta}_{z,n} = \theta_*$, \mathbb{P}^{θ_*} -a.s., holds for any $z \in \mathbb{Z}$.

Before proving Proposition 15, let us show the following result.

LEMMA 16. *Assume (4.43). Then*

$$(4.44) \quad \tilde{\mathbb{E}}^{\theta_*}[Y_1^2] < \infty.$$

PROOF. Now, assume that (4.43) holds. To obtain (4.44) it is sufficient to show that there exists a triplet $(q, \lambda, \beta) \in \mathbb{Z}_+ \times (0, 1) \times \mathbb{R}_+$ such that $R^q \tilde{V} < \lambda \tilde{V} + \beta$ with $\tilde{V}(z) = z^T z$. Note first that for $U \sim \mathcal{NB}(r, x/(1+x))$, we have $\mathbb{E}[U] = rx$ and $\mathbb{E}[U^2] = r(r+1)x^2 + rx$. Note also that the matrix $\mathbf{M} := \mathbf{A}^T \mathbf{A} + r (\mathbf{A}^T \mathbf{b} \boldsymbol{\delta}_p^T + \boldsymbol{\delta}_p \mathbf{b}^T \mathbf{A}) + r(r+1) \boldsymbol{\delta}_p \mathbf{b}^T \mathbf{b} \boldsymbol{\delta}_p^T$ is symmetric positive semidefinite. Thus, there exist a diagonal matrix $\mathbf{D} \geq 0$ and an orthogonal matrix \mathbf{Q} such that $\mathbf{M} = \mathbf{Q} \mathbf{D} \mathbf{Q}^T$; and, moreover, $|\lambda|_{\max}(\mathbf{M}) = |\lambda|_{\max}(\mathbf{D})$.

For all $z \in \mathbb{Z}$, we have

$$\begin{aligned} R\tilde{V}(z) &= \int \tilde{V}(\boldsymbol{\omega} + \mathbf{A}z + y\mathbf{b}) G(\Pi_p(z); dy) \\ &= \int (\boldsymbol{\omega} + \mathbf{A}z + y\mathbf{b})^T (\boldsymbol{\omega} + \mathbf{A}z + y\mathbf{b}) G(\boldsymbol{\delta}_p^T z; dy) \\ &= L_1(z) + z^T \mathbf{M} z = L_1(z) + z^T \mathbf{Q} \mathbf{D} \mathbf{Q}^T z \\ &\leq L_1(z) + |\lambda|_{\max}(\mathbf{D}) z^T \mathbf{Q} \mathbf{Q}^T z = L_1(z) + |\lambda|_{\max}(\mathbf{M}) \tilde{V}(z), \end{aligned}$$

where $L_1(z) = \boldsymbol{\omega}^T \boldsymbol{\omega} + (2\boldsymbol{\omega}^T \mathbf{A} + r2\boldsymbol{\omega}^T \mathbf{b}\boldsymbol{\delta}_p^T + r\mathbf{b}^T \mathbf{b}\boldsymbol{\delta}_p^T) z$, which is a linear function in z . Since $|\lambda|_{\max}(\mathbf{M}) < 1$, and

$$\lim_{|z| \rightarrow \infty} \frac{R\tilde{V}(z)}{\tilde{V}(z)} \leq |\lambda|_{\max}(\mathbf{M}) < 1,$$

then there exists $(\beta, \lambda) \in \mathbb{R}_+ \times (0, 1)$ such that for all $z \in \mathbb{Z}$,

$$R\tilde{V}(z) \leq \lambda\tilde{V}(z) + \beta,$$

completing the proof. \square

PROOF OF PROPOSITION 15. Assumption (AG-1) holds by Remark 12. We now show that (CG-1)–(CG-7) are satisfied. From the definition of $\Psi_y^\theta(z)$, we have $a(\theta, y) = \mathbf{A}$ and $b(\theta, y) = \boldsymbol{\omega} + y\mathbf{b}$, both admitting linear forms in θ . From compactness of Θ and by noting that (4.43) implies (CG-1)(ii) and that (4.44) also implies

$$(4.45) \quad \tilde{\mathbb{E}}^{\theta_\star} [Y_1] < \infty,$$

it follows that (CG-1) holds. Assumptions (CG-2) and (CG-3) are clearly satisfied. Assumption (CG-4) immediately follows by Remark 12. We have

$$\begin{aligned} \ell_{z,k}^\theta &= \ln g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{1:k-1} \rangle(z)); Y_k \right) \\ &= \ln \Gamma(Y_k + r) - (Y_k + r) \ln \left(1 + \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right) \\ &\quad + Y_k \ln \left(\boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right) - \ln \Gamma(r) - \ln(Y_k!), \end{aligned}$$

$$\begin{aligned} \ell_k^\theta &= \ln g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:k-1} \rangle); Y_k \right) \\ &= \ln \Gamma(Y_k + r) - (Y_k + r) \ln \left(1 + \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle \right) \\ &\quad + Y_k \ln \left(\boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle \right) - \ln \Gamma(r) - \ln(Y_k!). \end{aligned}$$

Then from Lemma 5, for all $\theta = \theta_{0:2p+1}$ with $\theta_{2p+1} = r$, for all $i \neq 2p+1$, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\begin{aligned} \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} &= \left(\frac{Y_k}{\boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z)} - \frac{Y_k + r}{1 + \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z)} \right) \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{1:k-1} \rangle(z) \\ \frac{\partial \ell_{z,k}^\theta}{\partial \theta_{2p+1}} &= \Gamma_2(Y_k + r) - \Gamma_2(r) - \ln \left(1 + \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \ell_k^\theta}{\partial \theta_i} &= \left(\frac{Y_k}{\delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle} - \frac{Y_k + r}{1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle} \right) \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle \\ \frac{\partial \ell_k^\theta}{\partial \theta_{2p+1}} &= \Gamma_2(Y_k + r) - \Gamma_2(r) - \ln \left(1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle \right), \end{aligned}$$

where Γ_2 is the digamma function defined by, for all $u > 0$,

$$\Gamma_2(u) := \frac{d}{du} \ln \Gamma(u) = \frac{\int_0^\infty e^{-t} t^{u-1} \ln t \, dt}{\int_0^\infty e^{-t} t^{u-1} \, dt}.$$

For $u \geq 3$, the digamma function satisfies

$$\begin{aligned} |\Gamma_2(u)| &\leq \frac{\int_0^\infty e^{-t} t^{u-1} |\ln t| \, dt}{\Gamma(u)} \leq \frac{\int_0^\infty e^{-t} t^{u-1} (t + t^{-1}) \, dt}{\Gamma(u)} \\ (4.46) \quad &= \frac{\Gamma(u+1) + \Gamma(u-1)}{\Gamma(u)} = u + (u-1)^{-1} \leq u+1, \end{aligned}$$

where we have used that $|\ln t| \leq (t + t^{-1})$ whenever $t > 0$.

It is straight forward to have that, for all i , $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned} \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| &\leq \left| \delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right| \\ &+ \left| \frac{Y_k + r}{1 + \delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z)} - \frac{Y_k}{\delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z)} - \right. \\ &\left. \frac{Y_k + r}{1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle} + \frac{Y_k}{\delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle} \right| \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle \right| \\ &+ \left| \frac{Y_k + r}{1 + \delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z)} - \frac{Y_k}{\delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z)} \right| \\ &\times \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right|. \end{aligned}$$

Since $\delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \geq \underline{\omega}$ and so that $\delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle \geq \underline{\omega}$, we obtain that for all i , $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned}
& \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| \leq \left| \delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right| \\
& + \left(\frac{2Y_k + r}{\underline{\omega}} \right) \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle \right| \left| \delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right| \\
& + \left(\frac{2Y_k + r}{\underline{\omega}} \right) \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right|.
\end{aligned}$$

Note that for all $z \in \mathbf{Z}$ and $k \geq 1$, we have, $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\Psi^\theta \langle Y_{-\infty:k-1} \rangle - \Psi^\theta \langle Y_{1:k-1} \rangle(z) = \mathbf{A}^{k-1} \left(\Psi^\theta \langle Y_{-\infty:0} \rangle - z \right),$$

$$\frac{\partial}{\partial \theta_{2p+1}} \Psi^\theta \langle Y_{1:k-1} \rangle(z) = \frac{\partial}{\partial \theta_{2p+1}} \Psi^\theta \langle Y_{-\infty:k-1} \rangle = 0 \text{ and, for all } i \neq 2p+1,$$

$$\begin{aligned}
& \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{1:k-1} \rangle(z) \\
& = \frac{\partial (\mathbf{A}^{k-1})}{\partial \theta_i} \left(\Psi^\theta \langle Y_{-\infty:0} \rangle - z \right) + \mathbf{A}^{k-1} \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle.
\end{aligned}$$

By proceeding similarly as in the proof of [Lemma 5](#) and noting that $\sup_{\theta \in \Theta} |\delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle|$ and $\sup_{\theta \in \Theta} \left| \delta_p^T \frac{\partial}{\partial \theta} \Psi^\theta \langle Y_{-\infty:0} \rangle \right|$ are finite $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., then there exist a constant $\rho_1 \in (0, 1)$ and a $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. finite random variable $\tilde{M}_1 = \tilde{M}_1(z)$ such that for all $k \geq 1$, $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right| \leq \tilde{M}_1 \rho_1^k,$$

and for all i , $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right| \leq \tilde{M}_1 \rho_1^k.$$

The last two inequalities imply that, for all $\theta \in \Theta$, $i \in \{1, \dots, 2p+2\}$ and $k \in \mathbb{Z}_+^*$, $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| \leq \tilde{M}_1 \rho_1^k \left[1 + \left(\frac{2Y_k + \bar{r}}{\underline{\omega}} \right) \left(1 + \sup_{\theta \in \Theta} \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle \right| \right) \right].$$

Note also that from [Lemma 21](#), we have, for all i ,

$$(4.47) \quad \tilde{\mathbb{E}}^{\theta^*} \left[\sup_{\theta \in \Theta} \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle \right| \right] < \infty.$$

Thus, by stationarity and from (4.45) and (4.47), Lemma 21 implies that there exist a constant $\rho_2 \in (0, 1)$ and a $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. finite random variable $\tilde{M}_2 = \tilde{M}_2(z)$ such that for all $\theta \in \Theta$, $i \in \{0, \dots, 2p+1\}$ and $k \geq 1$, $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| \leq \rho_2^k \tilde{M}_2.$$

Therefore, Assumption (CG-5) holds. We now check (CG-6). We have

$$\begin{aligned} \ell_1^\theta &= \ln g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1 \right) \\ &= \ln \Gamma(Y_1 + r) - (Y_1 + r) \ln \left(1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle \right) \\ &\quad - Y_1 \ln \left(\delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle \right) - \ln \Gamma(r) - \ln(Y_1!). \end{aligned}$$

Write $\theta = \theta_{0:2p+1}$. Then for all i and j such that $i \neq 2p+1$ and $j \neq 2p+1$, we have

$$(4.48) \quad \frac{\partial \ell_1^\theta}{\partial \theta_i} = \left(\frac{Y_1}{\delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle} - \frac{Y_1 + r}{1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle} \right) \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle,$$

$$(4.49) \quad \frac{\partial \ell_1^\theta}{\partial \theta_{2p+2}} = \Gamma_2(Y_1 + r) - \Gamma_2(r) - \ln \left(1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle \right),$$

and

$$\begin{aligned} \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_j} &= \left(\frac{Y_1}{\delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle} - \frac{Y_1 + r}{1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle} \right) \delta_p^T \frac{\partial^2}{\partial \theta_i \partial \theta_j} \Psi^\theta \langle Y_{-\infty:0} \rangle + \\ &\quad \left(\frac{Y_1 + r}{(1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle)^2} - \frac{Y_1}{(\delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle)^2} \right) \delta_p^T \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta_i} \delta_p^T \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta_j}, \\ \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_{2p+1}} &= - \left(1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle \right)^{-1} \delta_p^T \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta_i}, \\ \frac{\partial^2 \ell_1^\theta}{\partial \theta_{2p+1} \partial \theta_{2p+1}} &= \Gamma_3(Y_1 + r) - \Gamma_3(r), \end{aligned}$$

where Γ_3 is the trigamma function defined by, for all $u > 0$,

$$\Gamma_3(u) := \frac{d}{du} \Gamma_2(u) = \frac{\int_0^\infty e^{-t} t^{u-1} |\ln t|^2 dt}{\Gamma(u)} - \Gamma_2(u)^2.$$

Note that when $u \geq 3$, the trigamma function satisfies

$$\begin{aligned} |\Gamma_3(u)| &\leq \frac{\int_0^\infty e^{-t} t^{u-1} (t + t^{-1})^2 dt}{\Gamma(u)} + \Gamma_2(u)^2 \\ (4.50) \quad &= \frac{\Gamma(u+2) + 2\Gamma(u) + \Gamma(u-2)}{\Gamma(u)} + \Gamma_2(u)^2 \leq 2(u+1)^2 + 3. \end{aligned}$$

Using similar lines as the proof of [Lemma 5](#) together with [\(4.44\)](#), it can be shown that, for all i and j ,

$$(4.51) \quad \tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle \right|^2 \right] < \infty$$

and

$$(4.52) \quad \tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \Theta} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \Psi^\theta \langle Y_{-\infty:0} \rangle \right| \right] < \infty.$$

From [\(4.48\)](#), [\(4.49\)](#) and [\(4.46\)](#), we have, for all $i \neq 2p+1$, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\begin{aligned} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right| &\leq \left(\frac{2Y_k + \bar{r}}{\underline{\omega}} \right) \sup_{\theta \in \Theta} \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle \right|, \\ \left| \frac{\partial \ell_1^\theta}{\partial \theta_{2p+2}} \right| &\leq Y_1 + \tilde{C} + \ln \left(1 + \sup_{\theta \in \Theta} \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle \right| \right), \end{aligned}$$

where \tilde{C} is a positive constant. Using tower property, [\(4.44\)](#) and [\(4.51\)](#), we obtain that for all i ,

$$\tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right|^2 \right] < \infty.$$

By similar argument, we can obtain from [\(4.44\)](#), [\(4.50\)](#), [\(4.51\)](#) and [\(4.52\)](#) that, for all i and j ,

$$\tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \Theta} \left| \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_j} \right| \right] < \infty.$$

Thus, [\(CG-6\)-\(i\)](#) and [\(CG-6\)-\(ii\)](#) hold. Note that for all i, j , we have, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\frac{\partial}{\partial \theta_i} g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1 \right) = g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1 \right) \frac{\partial \ell_1^\theta}{\partial \theta_i}$$

and

$$\begin{aligned} &\frac{\partial^2}{\partial \theta_i \partial \theta_j} g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1 \right) \\ &= g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1 \right) \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_j} + \frac{\partial}{\partial \theta_i} g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1 \right) \frac{\partial \ell_1^\theta}{\partial \theta_j} \end{aligned}$$

Then, (CG-6)-(iii) follows by observing that for all i , $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \frac{\partial}{\partial \theta_i} g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1 \right) \right| = g^\theta \left(\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1 \right) \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right| \leq \sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right|$$

and

$$\tilde{\mathbb{E}}^{\theta^*} \left[\sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right| \right] \leq \left(\tilde{\mathbb{E}}^{\theta^*} \left[\sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right|^2 \right] \right)^{1/2} < \infty.$$

Similarly, (CG-6)-(iv) holds. For concluding, it remains to check (CG-7). By Remark 4, it is sufficient to show that (3.23) holds. Now let $\alpha = (\alpha_0, \dots, \alpha_{2p+1}) \in \mathbb{R}^{2p+2}$ and for convenience, write $\theta = (\theta_0, \dots, \theta_{2p+1}) = (\omega, a_{1:p}, b_{2:p}, b_1, r)$. We have $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned} \alpha^T \frac{\partial \ell_1^{\theta^*}}{\partial \theta} &= \left(\frac{Y_1}{\delta_p^T \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle} - \frac{Y_1 + r^*}{1 + \delta_p^T \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle} \right) \sum_{i=0}^{2p} \alpha_i \delta_p^T \frac{\partial \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle}{\partial \theta_i} \\ &\quad + \alpha_{2p+1} \left[\Gamma_2(Y_1 + r^*) - \Gamma_2(r^*) - \ln \left(1 + \delta_p^T \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle \right) \right]. \end{aligned}$$

Since Y_1 conditionally on $\sigma(Y_{-\infty:0})$ is a negative binomial variable, then $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., $\alpha^T \frac{\partial \ell_1^{\theta^*}}{\partial \theta} = 0$ implies

$$\alpha_{2p+1} \left[\Gamma_2(r^*) + \ln \left(1 + \delta_p^T \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle \right) \right] = 0, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

Since $X_1^{\theta^*} = \delta_p^T Z_1^{\theta^*} = \delta_p^T \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle$ is a random variable, then $\alpha_{2p+1} = 0$. Thus, $\alpha^T \frac{\partial \ell_1^{\theta^*}}{\partial \theta} = 0$, $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., also implies

$$\sum_{i=0}^{2p} \alpha_i \delta_p^T \frac{\partial \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle}{\partial \theta_i} = 0, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

The rest of the proof follows by the same argument as in the proof of Proposition 12-(iii) and is thus omitted here. The proof therefore follows. \square

5. Empirical Study.

5.1. *Sharpness of Ergodicity Condition (4.8) for the Log-linear Poisson GARCH Model.* We assess the sharpness of our general ergodicity condition

(4.8) for the log-linear Poisson GARCH model by simulating several time series with different values of the condition level

$$\nu_L := \bigvee_{(k_1, \dots, k_p) \in \{1, 2\}^p} \left\| \prod_{j=1}^p \mathbf{A}_{k_j} \right\|_{\infty}$$

and in the cases where $p = 1$ and 2 . Each generated time series holds 2000 observations. [Figure 1](#) presents three log-linear Poisson GARCH time series of order $(1, 1)$ with different parameters: Series ODM(1, 1)-(a) corresponds to the model with parameter $(\omega, a_1, b_1) = (-.03, -.7, 1.2)$ and $\nu_L = .7$; Series ODM(1, 1)-(b) corresponds to the model with parameter $(\omega, a_1, b_1) = (-.03, .57, .38)$ and $\nu_L = .95$; and Series ODM(1, 1)-(c) corresponds to the model with parameter $(\omega, a_1, b_1) = (-.03, .603, .402)$ and $\nu_L = 1.005$. Remark that in Series ODM(1, 1)-(a), we allow the parameter b_1 to be greater than one. [Figure 2](#) exhibits three log-linear Poisson GARCH time series of order $(2, 2)$ with different parameters. Series ODM(2, 2)-(a) corresponds to the model with parameter $(\omega, a_1, a_2, b_1, b_2) = (-.02, .4, .1, .3, .1)$ and $\nu_L = .9$; Series ODM(2, 2)-(b) corresponds to the model with parameter $(\omega, a_1, a_2, b_1, b_2) = (-.02, .44, .11, .33, .11)$ and $\nu_L = .99$; and Series ODM(2, 2)-(c) corresponds to the model with parameter $(\omega, a_1, a_2, b_1, b_2) = (-.02, .446, .1115, .3345, .1115)$ and $\nu_L = 1.0062$. For both [Figure 1](#) and [Figure 2](#), it turns out that when the condition level ν_L is far away from its boundary ($\nu_L < 1$), the processes appear quite stable. However, the level of instability seems to increase as the condition level ν_L gets closer to one. When this boundary is passed, the models appear to be unstable. We should remark that when $\nu_L < 1$, the stable behavior of these processes is persistent regardless of their starting points.

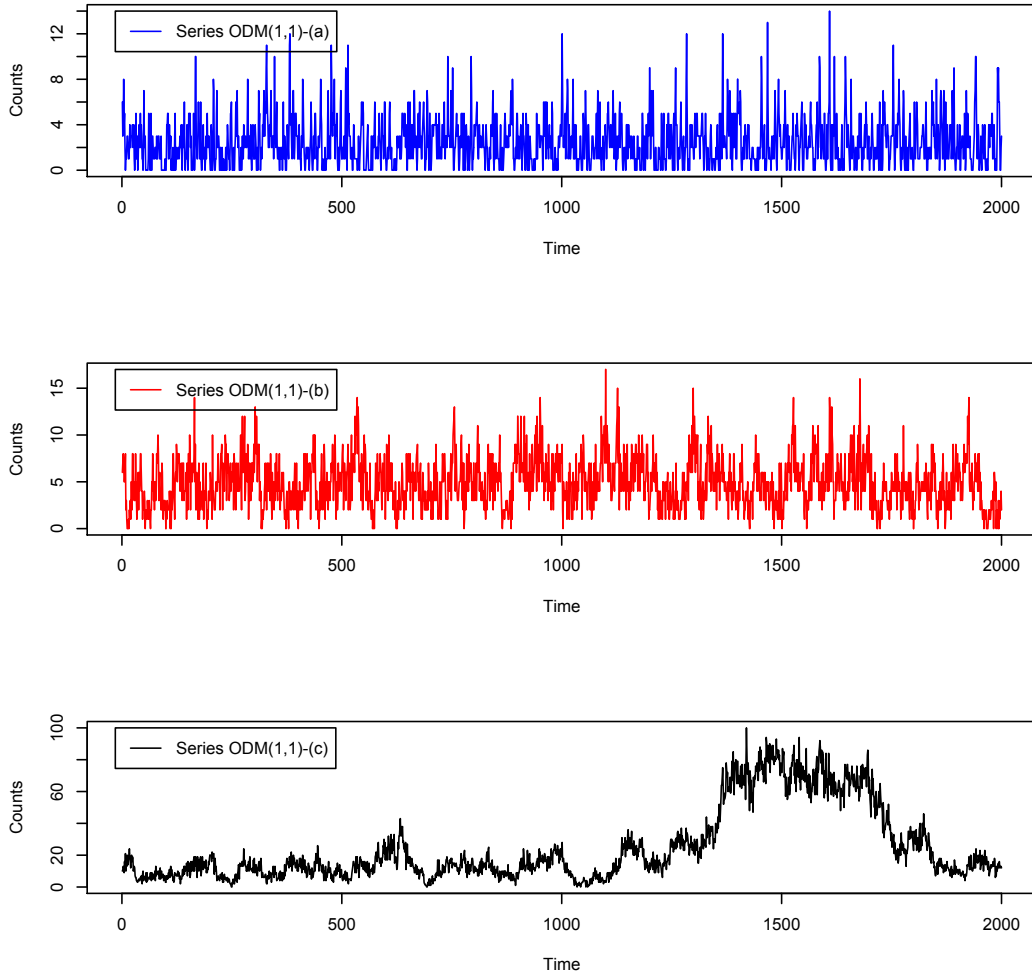


FIG 1. Simulated time series for the log-linear Poisson GARCH of order (1,1). Series ODM(1,1)-(a), -(b) and -(c) corresponds to the parameters (ω, a_1, b_1) equal to $(-.03, -.7, 1.2)$, $(-.03, .57, .38)$ and $(-.03, .603, .402)$, respectively.

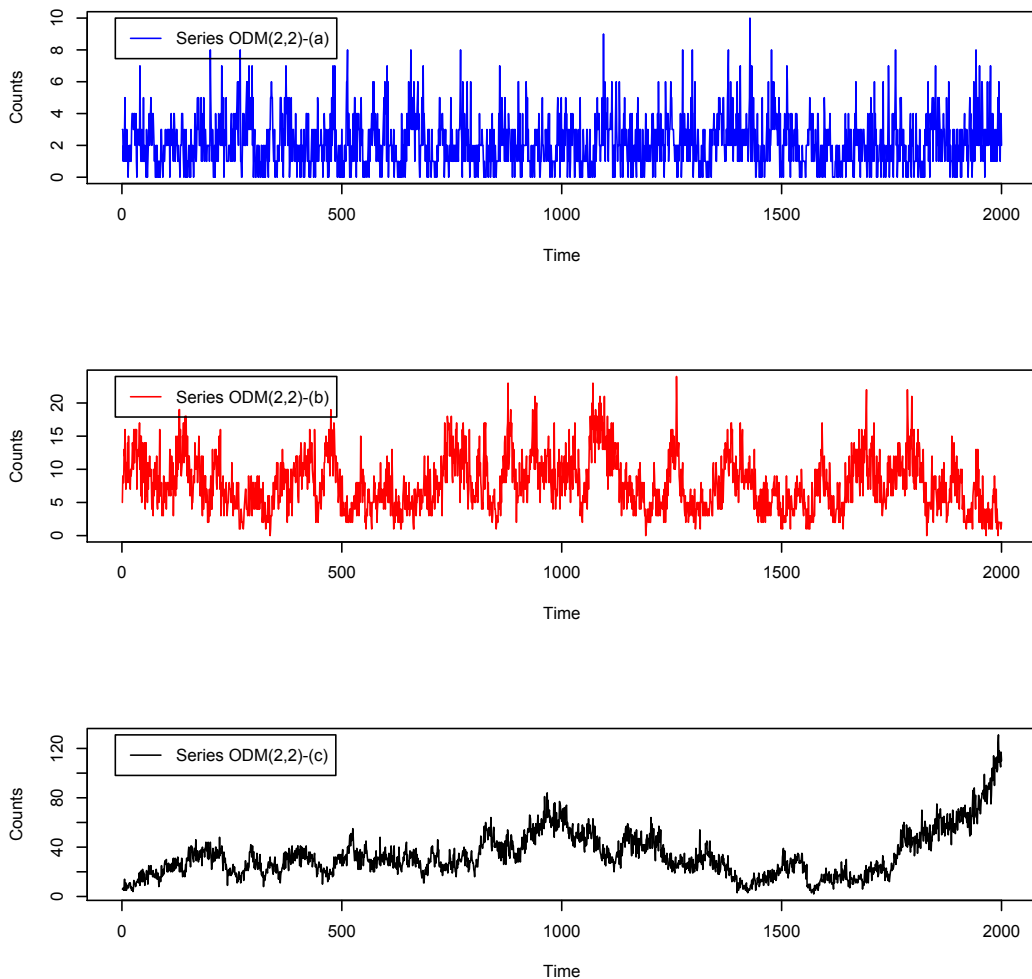


FIG 2. Simulated time series for the log-linear Poisson GARCH of order (2,2). Series ODM(2,2)-(a), -(b) and -(c) corresponds to the parameters $(\omega, a_1, a_2, b_1, b_2)$ equal to $(-.02, .4, .1, .3, .1)$, $(-.02, .44, .11, .33, .11)$ and $(-.02, .446, .1115, .3345, .1115)$, respectively.

5.2. *Data Example.* Our purpose here is to give some empirical evidence on some particular time series data for which higher-order ODMs may be more appropriate than the first-order ones. Our data set, provided by U.S. Geology Survey (USGS) (downloadable on: <http://earthquake>.

usgs.gov/earthquakes/search/), contains a time series of numbers of earthquakes occurring hourly around the globe with magnitudes larger than 4.0 Richters (a level that is detectable by humans), consisting of 8759 observations during the year 2014, from January 1 to December 31. The time series and its serial correlation plots are shown in [Figure 3](#) and [Figure 4](#), respectively, exhibiting that data possesses a long memory. In our analysis, we used the first 8559 observations for conducting the parameter estimation and the last 200 observations for evaluating the prediction performance of the fitted models. We fitted the data by several different models, by the negative binomial integer-valued GARCH (NBIN-GARCH) and the log-linear Poisson GARCH models with different orders: (1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2) and (3, 3). For estimating the parameters, we use the constrained nonlinear optimization function `auglag` (Augmented Lagrangian Minimization Algorithm) from the package `alabama` (Augmented Lagrangian Adaptive Barrier Minimization Algorithm) in R. The values of AIC, BIC and average squared prediction error (ASPE), $ASPE = m^{-1} \sum_{k=1}^m (Y_k - \hat{Y}_k)^2$, where \hat{Y}_k is the predictor of the observation Y_k —hence the mean process evaluated at the MLE, for these models are reported in [Table 1](#). It shows that, using either the AIC or BIC criterion, the higher-order models are preferred for both the NBIN-GARCH and the log-linear Poisson GARCH, opposed to their first-order counterparts. Among the fitted models, for the NBIN-GARCH, the AIC suggests the model of order (3, 3) corresponding to the estimator $(\hat{\omega}, \hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{b}_1, \hat{b}_2, \hat{b}_3, \hat{r}) = (.035, .424, .068, .222, .021, 1.036 \times 10^{-8}, 5.681 \times 10^{-8}, 7.431)$, while the BIC favors the model of order (3, 1) corresponding to the estimated parameter $(\hat{\omega}, \hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{b}_1, \hat{r}) = (.036, .414, .067, .232, .022, 7.131)$. We observe that the values of \hat{b}_2 and \hat{b}_3 are so small that they might be insignificant to be included. However, to discard these coefficients from the model properly, we may need to conduct a hypothesis testing and this is beyond the scope of this paper. Meanwhile, let us keep these terms in the model for convenience. For the log-linear Poisson GARCH models, the AIC suggests the model of (3, 3) corresponding to the estimator $(\hat{\omega}, \hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{b}_1, \hat{b}_2, \hat{b}_3) = (-.043, .495, .142, .091, .256, 1.084 \times 10^{-5}, -.017)$, whereas the BIC picks the model with order (2, 1) corresponding to the estimator $(\hat{\omega}, \hat{a}_1, \hat{a}_2, \hat{b}_1) = (-.041, .471, .234, .251)$. The estimators for the NBIN-GARCH(1, 1) and for the log-linear Poisson GARCH(1, 1) are $(\hat{\omega}, \hat{a}_1, \hat{b}_1, \hat{r}) = (.041, .719, .019, 6.997)$ and $(\hat{\omega}, \hat{a}_1, \hat{b}_1) = (-.028, .722, .226)$, respectively. [Table 1](#) also shows that the ASPE values of the models selected by either the AIC or the BIC criterion are lower than those of the first-order ones. This suggests that the higher-order models outperform in both cases of the NBIN-

GARCH and the log-linear Poisson GARCH. However, there are barely any differences between these values for the models selected by the AIC and by the BIC, respectively. By comparing across the models, the selected log-linear Poisson GARCH model seems to yield prediction performance as well as the selected NBIN-GARCH model in our data. [Figure 5](#) and [Figure 6](#) respectively depict the predicted series of first-order models and the ones chosen by AIC and BIC, compared with the real observations, for the NBIN-GARCH and the log-linear Poisson GARCH.

TABLE 1

AIC, BIC and ASPE values for the NBIN-GARCH and the log-linear Poisson GARCH models with different orders: (1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2) and (3, 3).

	NBIN-GARCH						
	(1, 1)	(1, 2)	(2, 1)	(2, 2)	(3, 1)	(3, 2)	(3, 3)
AIC	30102	30097	30083	30078	30074	30070	30065
BIC	30130	30132	30118	30120	30117	30119	30122
ASPE	1.633	1.633	1.618	1.633	1.614	1.614	1.613
	Log-linear Poisson GARCH						
	(1, 1)	(1, 2)	(2, 1)	(2, 2)	(3, 1)	(3, 2)	(3, 3)
AIC	30632	30621	30616	30614	30610	30608	30607
BIC	30653	30649	30644.5	30649	30645.4	30651	30656
ASPE	1.654	1.658	1.616	1.616	1.613	1.615	1.617

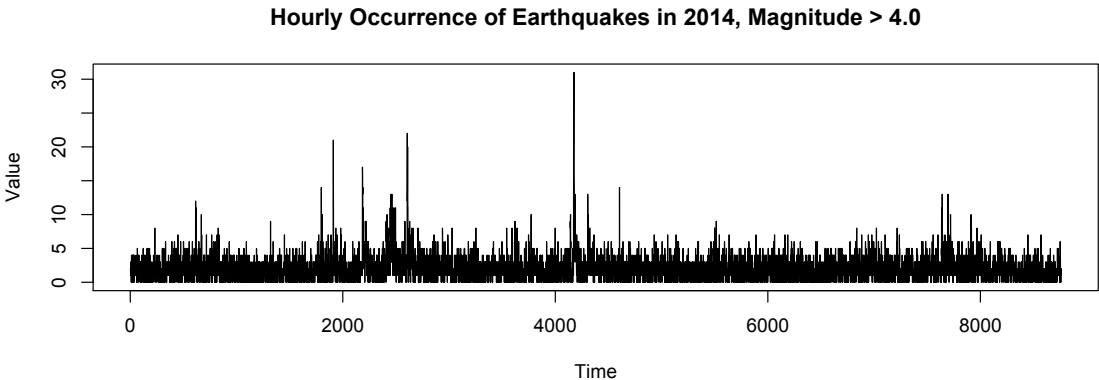


FIG 3. *Time series of numbers of earthquakes in 2014, from January 1 to December 31 (hourly).*

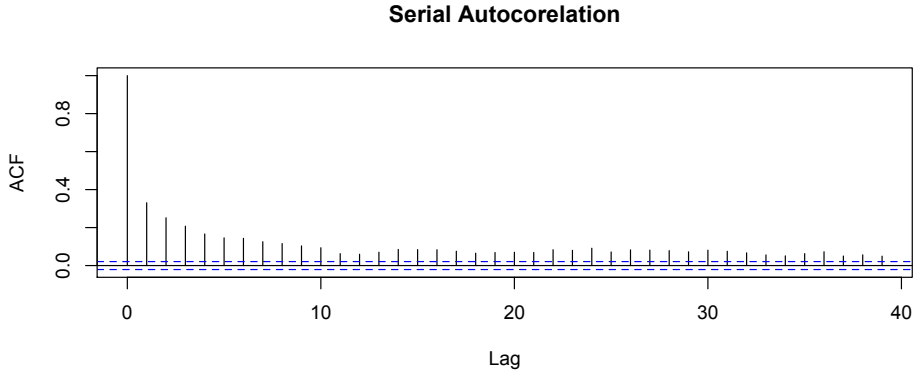


FIG 4. *Sample autocorrelation function.*

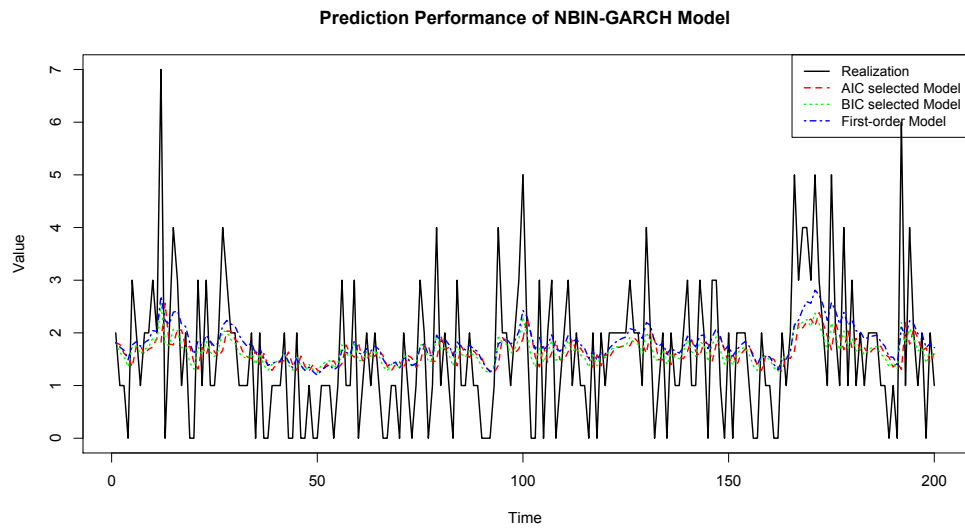


FIG 5. *Predicted series by the NBIN-GARCH of order (1,1) (dashed blue), by the one selected by the AIC (long dashed red) and by the one selected by the BIC (small dotted green), compared to the true series (black).*

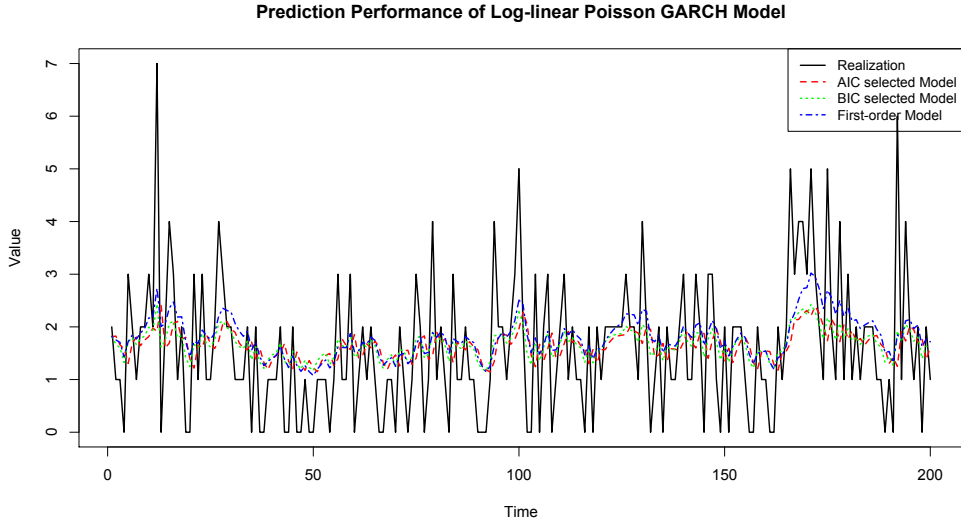


FIG 6. Predicted series by the log-linear Poisson GARCH of order $(1, 1)$ (dashed blue), by the one selected by the AIC (long dashed red) and by the one selected by the BIC (small dotted green), compared to the true series (black).

6. Postponed Proofs.

6.1. Proof of Lemma 5.

PROOF. From (3.13), for all $n \in \mathbb{Z}_+^*$, $\theta \in \Theta$, $z \in \mathbb{Z}$, and $y_{1:n} \in \mathbb{Y}^n$, we have

$$(6.1) \quad \Psi^\theta \langle y_{1:n} \rangle (z) = \left(\prod_{\ell=1}^n a(\theta, y_{n-\ell+1}) \right) z + \sum_{k=1}^n \left(\prod_{\ell=1}^{k-1} a(\theta, y_{n-\ell}) \right) b(\theta, y_{n-k+1}).$$

Then, for all $m \in \mathbb{Z}_+$, $\theta \in \Theta$ and $z \in \mathbb{Z}$, we get that

$$\Psi^\theta \langle Y_{-m:0} \rangle (z) = \left(\prod_{\ell=0}^m a(\theta, Y_{-\ell}) \right) z + \sum_{k=0}^m \left(\prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}).$$

By (CG-1)(ii), we have

$$\lim_{m \rightarrow \infty} \sup_{\theta \in \Theta} \left\| \left(\prod_{\ell=0}^m a(\theta, Y_{-\ell}) \right) z \right\| = 0, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

Moreover, we have

$$\sum_{k=0}^{\infty} \left\| \left(\prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \right\| \leq \bar{C} \sum_{k=0}^{\infty} \rho^k \sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\|,$$

for some positive constant \bar{C} . By (CG-1)(iii) and Lemma 21, the right-hand side of the above inequality converges $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. This implies that for all $z \in \mathbb{Z}$, the sequence $\{\Psi^\theta \langle Y_{-m:0} \rangle(z) : m \in \mathbb{Z}_+\}$ converges uniformly on Θ to a limit $\Psi^\theta \langle Y_{-\infty:0} \rangle := \sum_{k=0}^{\infty} \left(\prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k})$, $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s., showing that (3.4) is satisfied. Next we show that (3.5) is satisfied. Note that for all $m \in \mathbb{Z}_+$

$$\Psi^{\theta_\star} \langle Y_{-m:0} \rangle(Z_{-m}) = \left(\prod_{\ell=0}^m a(\theta_\star, Y_{-\ell}) \right) Z_{-m} + \sum_{k=0}^m \left(\prod_{\ell=0}^{k-1} a(\theta_\star, Y_{-\ell}) \right) b(\theta_\star, Y_{-k}).$$

Since $\{Z_{-m} : m \in \mathbb{Z}_+\}$ is stationary under $\mathbb{P}^{\theta_\star}$ so that it is bounded in probability, and since $\lim_{m \rightarrow \infty} \prod_{\ell=0}^m a(\theta_\star, Y_{-\ell}) = 0$, $\mathbb{P}^{\theta_\star}$ -a.s., it then implies that $\lim_{m \rightarrow \infty} \left(\prod_{\ell=0}^m a(\theta_\star, Y_{-\ell}) \right) Z_{-m} = 0$ in $\mathbb{P}^{\theta_\star}$ -probability. From the proof above, we have

$$\lim_{m \rightarrow \infty} \sum_{k=0}^m \left(\prod_{\ell=0}^{k-1} a(\theta_\star, Y_{-\ell}) \right) b(\theta_\star, Y_{-k}) = \Psi^{\theta_\star} \langle Y_{-\infty:0} \rangle, \quad \mathbb{P}^{\theta_\star}\text{-a.s.}$$

Note also that for all $m \in \mathbb{Z}_+$, $Z_1 = \Psi^{\theta_\star} \langle Y_{-m:0} \rangle(Z_{-m})$, $\mathbb{P}^{\theta_\star}$ -a.s. We therefore get that

$$\Psi^{\theta_\star} \langle Y_{-\infty:0} \rangle = Z_1, \quad \mathbb{P}^{\theta_\star}\text{-a.s.},$$

showing (3.5). In addition, (CG-1)(iii) implies

$$\begin{aligned} \tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \Theta} \left\| \Psi^\theta \langle Y_{-\infty:0} \rangle \right\| \right] &\leq \bar{C} \sum_{k=0}^{\infty} \rho^k \tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\| \right] \\ &\leq \frac{\bar{C}}{1-\rho} \tilde{\mathbb{E}}^{\theta_\star} \left[\sup_{\theta \in \Theta} \|b(\theta, Y_1)\| \right] < \infty, \end{aligned}$$

showing (3.16).

For any i , we have

$$\begin{aligned}
& \frac{\partial \Psi^\theta \langle Y_{-m:0} \rangle (z)}{\partial \theta_i} \\
&= \frac{\partial [(\prod_{\ell=0}^m a(\theta, Y_{-\ell})) z]}{\partial \theta_i} + \sum_{k=0}^m \frac{\partial \left[\left(\prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \right]}{\partial \theta_i} \\
&= \sum_{\ell'=0}^m \left(\prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right) z + \sum_{k=0}^m \left(\sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',i}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \\
&+ \sum_{k=0}^m \left(\prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i},
\end{aligned}$$

where for all i, ℓ, ℓ' ,

$$(6.2) \quad a_{\ell',i}(\theta, Y_{-\ell}) = \begin{cases} a(\theta, Y_{-\ell}) & \text{if } \ell \neq \ell' \\ \frac{\partial}{\partial \theta_i} a(\theta, Y_{-\ell}) & \text{if } \ell = \ell'. \end{cases}$$

From (CG-1)(ii), we have

$$\begin{aligned}
& \left\| \sum_{\ell'=0}^m \left(\prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right) z \right\| \leq \sum_{\ell'=0}^m \left\| \prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right\| \|z\| \\
& \leq \sum_{\ell'=0}^m C^2 \rho^m \left\| \frac{\partial a(\theta, Y_{-\ell'})}{\partial \theta_i} \right\| \|z\| \\
& \leq C^2 \|z\| \left(\frac{1+\rho}{2} \right)^m \sum_{\ell'=0}^m \left(\frac{2\rho}{1+\rho} \right)^{\ell'} \sup_{\theta \in \Theta} \left\| \frac{\partial a(\theta, Y_{-\ell'})}{\partial \theta_i} \right\|.
\end{aligned}$$

We have from (CG-1)(iv) and by Lemma 21,

$$\sum_{\ell'=0}^{\infty} \left(\frac{2\rho}{1+\rho} \right)^{\ell'} \sup_{\theta \in \Theta} \left\| \frac{\partial a(\theta, Y_{-\ell'})}{\partial \theta_i} \right\| < \infty, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

Thus, $\tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$, $\left\| \sum_{\ell'=0}^m \left(\prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right) z \right\|$ converges uniformly (in θ) to zero. Moreover, there exist a constant $\rho_1 \in (0, 1)$ and a $\tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$ finite random variable $\tilde{M}_1 > 0$ with $\tilde{\mathbb{E}}^{\theta^*}[\tilde{M}_1] < \infty$ such that for all $m \in \mathbb{Z}_+$,

$$(6.3) \quad \sum_{\ell'=0}^m \sup_{\theta \in \Theta} \left\| \prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right\| \leq \tilde{M}_1 \rho_1^m, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

Next, we show that $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\sum_{k=0}^m \left(\sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',i}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k})$$

converges uniformly on Θ . We obtain from (6.3), $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\begin{aligned} \sum_{k=0}^m \left\| \left(\sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',i}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \right\| &\leq \tilde{M}_1 \sum_{k=0}^m \rho_1^{k-1} \sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\| \\ &\leq \tilde{M}_1 \sum_{k=0}^{\infty} \rho_1^{k-1} \sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\|, \end{aligned}$$

which is finite from (CG-1)(iii) and Lemma 21. To conclude (3.14), it remains to show that $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\sum_{k=0}^m \left(\prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i}$$

converges uniformly on Θ . We have $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\begin{aligned} \sum_{k=0}^m \left\| \left(\prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i} \right\| &\leq \sum_{k=0}^m C \rho^k \sup_{\theta \in \Theta} \left\| \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i} \right\| \\ &\leq C \sum_{k=0}^{\infty} \rho^k \sup_{\theta \in \Theta} \left\| \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i} \right\|, \end{aligned}$$

which is finite by (CG-1)(iv) and Lemma 21. Thus the sequence $\left\{ \frac{\partial \Psi^\theta \langle Y_{-m:0} \rangle (z)}{\partial \theta} : m \in \mathbb{Z}_+ \right\}$ $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. converges uniformly on Θ and this follows that for all $z \in \mathbb{Z}$ and $\theta \in \Theta$,

$$\lim_{m \rightarrow \infty} \frac{\partial \Psi^\theta \langle Y_{-m:0} \rangle (z)}{\partial \theta} = \frac{\partial}{\partial \theta} \lim_{m \rightarrow \infty} \Psi^\theta \langle Y_{-m:0} \rangle (z) = \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta}, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

showing that the quantity $\frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta}$ is $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. well defined. The fact that (3.17) holds follows from (CG-1)(iv), (3.16) and (6.3).

For any i, j , we have

$$\begin{aligned}
& \frac{\partial^2 \Psi^\theta \langle Y_{-m:0} \rangle (z)}{\partial \theta_j \partial \theta_i} \\
&= \sum_{\ell'=0}^m \frac{\partial}{\partial \theta_j} \left(\prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right) z \\
&+ \sum_{k=0}^m \frac{\partial}{\partial \theta_j} \left[\left(\sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',i}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \right] \\
&+ \sum_{k=0}^m \frac{\partial}{\partial \theta_j} \left[\left(\prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i} \right] \\
&= U_{1,m}(\theta) + U_{2,m}(\theta) + U_{3,m}(\theta) + U_{4,m}(\theta) + U_{5,m}(\theta),
\end{aligned}$$

where

$$\begin{aligned}
U_{1,m}(\theta) &= \sum_{\ell'=0}^m \sum_{\ell''=0}^m \left(\prod_{\ell=0}^m a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right) z, \\
U_{2,m}(\theta) &= \sum_{k=0}^m \left(\sum_{\ell'=0}^{k-1} \sum_{\ell''=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}), \\
U_{3,m}(\theta) &= \sum_{k=0}^m \left(\sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',i}(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_j}, \\
U_{4,m}(\theta) &= \sum_{k=0}^m \left(\sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',j}(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i}, \\
U_{5,m}(\theta) &= \sum_{k=0}^m \left(\prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) \frac{\partial^2 b(\theta, Y_{-k})}{\partial \theta_j \partial \theta_i},
\end{aligned}$$

and with for all $i, j, \ell, \ell', \ell''$,

$$(6.4) \quad a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) = \begin{cases} a_{\ell',i}(\theta, Y_{-\ell}) & \text{if } \ell \neq \ell'' \\ \frac{\partial a_{\ell',i}(\theta, Y_{-\ell})}{\partial \theta_j} & \text{if } \ell = \ell'', \end{cases}$$

and from (6.2),

$$(6.5) \quad \frac{\partial a_{\ell',i}(\theta, Y_{-\ell})}{\partial \theta_j} = \begin{cases} \frac{\partial a_{\ell',i}(\theta, Y_{-\ell})}{\partial \theta_j} & \text{if } \ell \neq \ell' \\ \frac{\partial^2 a_{\ell',i}(\theta, Y_{-\ell})}{\partial \theta_j \partial \theta_i} & \text{if } \ell = \ell'. \end{cases}$$

To establish that $\left\{ \frac{\partial^2 \Psi^\theta \langle Y_{-m;0} \rangle (z)}{\partial \theta_j \partial \theta_i} : m \in \mathbb{Z}_+ \right\}$ is almost surely uniformly convergent on Θ , it requires proving that for all $i \in \{1, \dots, 5\}$, the sequences $\{U_{i,m}(\theta) : m \in \mathbb{Z}_+\}$ converges uniformly on Θ . We have

$$\begin{aligned} \|U_{1,m}(\theta)\| &= \left\| \sum_{\ell'=0}^m \sum_{\ell''=0}^m \left(\prod_{\ell=0}^m a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right) z \right\| \\ &\leq \sum_{\ell'=0}^m \sum_{\ell''=0}^m \left\| \prod_{\ell=0}^m a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right\| \|z\| \\ &\stackrel{(1)}{\leq} \tilde{M}_1^2 \|z\| \sum_{\ell'=0}^m \sum_{\ell''=0}^m \rho_1^m \sup_{\theta \in \Theta} \left\| \frac{\partial a_{\ell',i}(\theta, Y_{-\ell''})}{\partial \theta_j} \right\| \\ &\stackrel{(2)}{\leq} \tilde{M}_1^2 \|z\| \sum_{\ell''=0}^m \rho_1^m \left(m \sup_{\theta \in \Theta} \left\| \frac{\partial a(\theta, Y_{-\ell''})}{\partial \theta_j} \right\| + \sup_{\theta \in \Theta} \left\| \frac{\partial^2 a(\theta, Y_{-\ell''})}{\partial \theta_j \partial \theta_i} \right\| \right), \end{aligned}$$

where $\stackrel{(1)}{\leq}$ follows from (6.3) and the submultiplicativity of the matrix norm $\|\cdot\|$, and $\stackrel{(2)}{\leq}$ follows from (6.5). By assuming

$$\tilde{\mathbb{E}}^{\theta_*} \left[\ln^+ \sup_{\theta \in \Theta} \left\| \frac{\partial a(\theta, Y_1)}{\partial \theta_i} \right\| \right] < \infty$$

and

$$\tilde{\mathbb{E}}^{\theta_*} \left[\ln^+ \sup_{\theta \in \Theta} \left\| \frac{\partial^2 a(\theta, Y_1)}{\partial \theta_j \partial \theta_i} \right\| \right] < \infty,$$

then there exist a constant $\rho_2 \in (0, 1)$ and a $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. finite random variable $\tilde{M}_2 > 0$ with $\tilde{\mathbb{E}}^{\theta_*}[\tilde{M}_2] < \infty$ such that for all $m \in \mathbb{Z}_+$,

$$\leq \sum_{\ell'=0}^m \sum_{\ell''=0}^m \left\| \prod_{\ell=0}^m a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right\| \leq \tilde{M}_2 \rho_2^m, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

Consequently, for all $z \in \mathbb{Z}$,

$$\lim_{m \rightarrow \infty} \sup_{\theta \in \Theta} \|U_{1,m}(\theta)\| = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

Thus, $\tilde{\mathbb{P}}^{\theta_*}$ -a.s., the sequence $\{U_{1,m}(\theta) : m \in \mathbb{Z}_+\}$ converges to zero uniformly on Θ . Next, we have

$$\begin{aligned}
\|U_{2,m}(\theta)\| &= \sum_{k=0}^m \left\| \left(\sum_{\ell'=0}^{k-1} \sum_{\ell''=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \right\| \\
&\leq \sum_{k=0}^m \left(\sum_{\ell'=0}^{k-1} \sum_{\ell''=0}^{k-1} \left\| \prod_{\ell=0}^{k-1} a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right\| \right) \|b(\theta, Y_{-k})\| \\
&\leq \tilde{M}_2 \sum_{k=0}^m \rho^{k-1} \sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\|.
\end{aligned}$$

Since $\lim_{m \rightarrow \infty} \sum_{k=0}^m \rho^{k-1} \sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\| < \infty$, then $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., the sequence $\{U_{2,m}(\theta) : m \in \mathbb{Z}_+\}$ converges uniformly on Θ . To show that $\{U_{3,m}(\theta) : m \in \mathbb{Z}_+\}$ and $\{U_{4,m}(\theta) : m \in \mathbb{Z}_+\}$ converge uniformly on Θ , $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., we proceed similarly as above and apply (6.3) and (CG-1)(iv). Finally, by the same argument, the almost surely uniform convergence of $\{U_{5,m}(\theta) : m \in \mathbb{Z}_+\}$ follows from (CG-1)(ii) and (CG-1)(v). The $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. uniform convergence of $\{\frac{\partial^2 \Psi^\theta \langle Y_{-m:0} \rangle(z)}{\partial \theta_j \partial \theta_i} : m \in \mathbb{Z}_+\}$ on Θ then implies that $\theta \mapsto \frac{\partial^2 \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta_j \partial \theta_i}$ is $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. well defined and

$$\frac{\partial^2 \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta_j \partial \theta_i} = \lim_{m \rightarrow \infty} \frac{\partial^2 \Psi^\theta \langle Y_{-m:0} \rangle(z)}{\partial \theta_j \partial \theta_i}, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.},$$

showing (3.15). The proof for (3.18) follows by similar arguments as for (3.16) and (3.17). \square

6.2. *Proof of Theorem 9.* The proof of this theorem follows similar lines of Douc, Roueff and Sim (2015). To proceed, let us first recall a more general set of conditions derived in Douc, Doukhan and Moulines (2013) in which the following definition is required.

DEFINITION 17. Let \bar{H} be a probability kernel from Z^2 to $\mathcal{Y}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\})$ satisfying the following marginal conditions, for all $(z, z') \in Z^2$ and $B \in \mathcal{Y}$,

$$(6.6) \quad \begin{cases} \bar{H}((z, z'); B \times \mathcal{Y} \times \{0, 1\}) = H(z; B) = G(\Pi_p(z); B), \\ \bar{H}((z, z'); \mathcal{Y} \times B \times \{0, 1\}) = H(z'; B) = G(\Pi_p(z'); B), \end{cases}$$

and such that the following coupling condition holds:

$$(6.7) \quad \bar{H}((z, z'); \{(y, y) : y \in \mathcal{Y}\} \times \{1\}) = \bar{H}((z, z'); \mathcal{Y}^2 \times \{1\}).$$

Define the following quantities successively.

- The trace measure of $\bar{H}((z, z'); \cdot)$ on the set $\{(y, y) : y \in \mathcal{Y}\} \times \{1\}$ is denoted by

$$\check{H}((z, z'); B) = \bar{H}((z, z'); \{(y, y) : y \in B\} \times \{1\}), \quad B \in \mathcal{Y}.$$

- The probability kernel \bar{R} from $(\mathbb{Z}^2, \mathcal{Z}^{\otimes 2})$ to $(\mathbb{Z}^2 \times \{0, 1\}, \mathcal{Z}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\}))$ is defined for all $z, z' \in \mathbb{Z}^2$ and $A \in \mathcal{X}^{\otimes 2}$ by

$$\bar{R}((z, z'); A \times \{1\}) = \int_{\mathcal{Y}} \mathbb{1}_A(\psi_y(z), \psi_y(z')) \check{H}((z, z'); dy).$$

- The measurable function α from \mathbb{Z}^2 to $[0, 1]$ is defined by

$$(6.8) \quad \alpha(z, z') = \bar{R}((z, z'); \mathbb{Z}^2 \times \{1\}) = \bar{H}((z, z'); \mathcal{Y}^2 \times \{1\}).$$

- The kernel \hat{R} is defined for all $(z, z') \in \mathbb{Z}^2$ and $A \in \mathcal{Z}^{\otimes 2}$ by

$$(6.9) \quad \hat{R}((z, z'); A) = \begin{cases} \frac{\bar{R}((z, z'); A \times \{1\})}{\alpha(z, z')} & \text{if } \alpha(z, z') > 0, \\ 0 & \text{otherwise.} \end{cases}$$

We can now introduce the so-called *contracting condition* which yields ergodicity.

- (AG-8) There exists a kernel \bar{H} yielding α and \hat{R} as in [Definition 17](#), a measurable function $W : \mathbb{Z}^2 \rightarrow [1, \infty)$ satisfying [Conditions \(AG-7\)\(ii\)](#) and [\(AG-7\)\(iii\)](#) and real numbers $(D, \zeta_1, \zeta_2, \rho) \in (\mathbb{R}_+)^3 \times (0, 1)$ such that for all $(z, z') \in \mathbb{X}^2$ and, for all $n \geq 1$,

$$(6.10) \quad \hat{R}^n((z, z'); dz) \leq D \rho^n dz(z, z'),$$

$$(6.11) \quad \hat{R}^n((z, z'); dz \times W) \leq D \rho^n d_Z^{\zeta_1}(z, z') W^{\zeta_2}(z, z').$$

Under [Conditions \(AG-3\)](#), [\(AG-4\)](#), [\(AG-5\)](#) and [\(AG-8\)](#) and by combining [Theorem 6](#), [Proposition 8](#) and [Lemma 7](#) in [Douc, Doukhan and Moulines \(2013\)](#), we immediately obtain the following result.

THEOREM 18. *Assume [\(AG-3\)](#), [\(AG-4\)](#), [\(AG-5\)](#) and [\(AG-8\)](#). Then the Markov kernel K admits a unique invariant distribution π and $\pi_1(\bar{V}) < \infty$ for any $\bar{V} : \mathbb{X} \rightarrow \mathbb{R}_+$ such that $\bar{V} \lesssim V$.*

Assumptions [\(AG-3\)](#), [\(AG-4\)](#) and [\(AG-5\)](#) are quite usual and easy to check. The key point to obtain ergodicity is thus to construct \bar{H} satisfying [\(AG-8\)](#). For this, we can also rely on the following result which is an adaption of [\(Douc, Doukhan and Moulines, 2013, Lemma 9\)](#).

LEMMA 19. Assume that there exists $(\varrho, \beta, D_1, D_2, \ell, \ell') \in (0, 1) \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{Z}_+^* \times \mathbb{Z}_+^*$ such that for all $(z, z') \in \mathbb{Z}^2$,

$$(6.12) \quad \hat{R}((z, z'); \mathbb{1}_{(D_1 d_{\mathbb{Z}}(z, z'), \infty)} \circ d_{\mathbb{Z}}) = 0,$$

$$(6.13) \quad \hat{R}^{\ell}((z, z'); \mathbb{1}_{(\varrho d_{\mathbb{Z}}(z, z'), \infty)} \circ d_{\mathbb{Z}}) = 0,$$

and

$$(6.14) \quad \hat{R}W \leq D_2 W,$$

$$(6.15) \quad \hat{R}^{\ell'} W \leq W + \beta.$$

Then, (6.10) and (6.11) hold.

REMARK 13. Lemma 19 still holds if we replace (6.14) and (6.15) by

$$\hat{R}^{\ell'+r} W \leq W + \beta, \quad r \in \{0, \dots, \ell' - 1\}.$$

PROOF. Without loss of generality, we can assume that $\ell = \ell' = m$, otherwise we can take $m = \ell \vee \ell'$. For positive integer n , write $n = km + r$, where $r \in \{0, \dots, m - 1\}$ and $k \in \mathbb{Z}_+$. We have

$$\begin{aligned} \hat{R}^n((z, z'); d_{\mathbb{Z}}) &= \hat{R}^{km+r}((z, z'); d_{\mathbb{Z}}) = \int \hat{R}^{km}((z, z'); dz_1 dz'_1) \hat{R}^r((z_1, z'_1); d_{\mathbb{Z}}) \\ &= \int \hat{R}^{km}((z, z'); dz_1 dz'_1) \hat{R}^r((z_1, z'_1); d_{\mathbb{Z}} \mathbb{1}_{[0, D_1 d_{\mathbb{Z}}(z_1, z'_1)]} \circ d_{\mathbb{Z}}) \\ &\leq \int \hat{R}^{km}((z, z'); dz_1 dz'_1) D_1^r d_{\mathbb{Z}}(z_1, z'_1) \\ &\leq D_1^r \int \hat{R}^{(k-1)m}((z, z'); dz_2 dz'_2) \hat{R}^m((z_2, z'_2); dz_1 dz'_1) d_{\mathbb{Z}}(z_1, z'_1) \\ &\leq \varrho D_1^r \int \hat{R}^{(k-1)m}((z, z'); dz_2 dz'_2) d_{\mathbb{Z}}(z_2, z'_2) \\ &\dots \\ &\leq \varrho^k D_1^r d_{\mathbb{Z}}(z, z') \leq \varrho^{-1} \left(\bigvee_{r=0}^{m-1} D_1^r \right) \left(\varrho^{1/m} \right)^n d_{\mathbb{Z}}(z, z'). \end{aligned}$$

Similarly, we have

$$\begin{aligned}
\hat{R}^n((z, z'); d_{\mathbf{Z}} \times W) &= \int \hat{R}^{km}((z, z'); dz_1 dz'_1) \hat{R}^r((z_1, z'_1); d_{\mathbf{Z}} \times W) \\
&\leq \int \hat{R}^{km}((z, z'); dz_1 dz'_1) D_1^r D_2^r d_{\mathbf{Z}}(z_1, z'_1) W(z_1, z'_1) \\
&\leq D_1^r D_2^r \int \hat{R}^{(k-1)m}((z, z'); dz_2 dz'_2) \hat{R}^m((z_2, z'_2); dz_1 dz'_1) d_{\mathbf{Z}}(z_1, z'_1) W(z_1, z'_1) \\
&\leq \varrho D_1^r D_2^r \int \hat{R}^{(k-1)m}((z, z'); dz_2 dz'_2) d_{\mathbf{Z}}(z_2, z'_2) (W(z_2, z'_2) + \beta) \\
&\dots \\
&\leq D_1^r D_2^r \varrho^k (W(z, z') + \beta k) d_{\mathbf{Z}}(z, z').
\end{aligned}$$

Since for all $\varrho \in (0, 1)$, $\lim_{k \rightarrow \infty} k \varrho^k = 0$, then there exists $\bar{D} > 0$ and $\bar{\varrho} \in [\varrho, 1)$ such that

$$\begin{aligned}
\hat{R}^n((z, z'); d_{\mathbf{Z}} \times W) &\leq \bar{D} \bar{\varrho}^{k-1} W(z, z') d_{\mathbf{Z}}(z, z') \\
&\leq \bar{D} (\bar{\varrho}^{1/m})^n W(z, z') d_{\mathbf{Z}}(z, z').
\end{aligned}$$

Setting $D = \bar{D} \vee \left(\varrho^{-1} \left(\bigvee_{r=0}^{m-1} D_1^r \right) \right)$ and $\rho = \varrho^{1/m} \vee \bar{\varrho}^{1/m}$, we get (6.10) and (6.11). \square

Now we can prove that our set of conditions is sufficient.

PROOF OF THEOREM 9. We only need to show that (AG-6) and (AG-7) imply (AG-8). We preface our proof by the following lemma, which is adapted from (Douc, Roueff and Sim, 2015, Lemma 5).

LEMMA 20. *Assume (AG-7)(i). Then one can define a kernel \bar{H} as in Definition 17 with the same α given in (6.8). Moreover, the kernel \hat{R} defined by (6.9) satisfies, for all $(z, z') \in \mathbb{Z}^2$ such that $\alpha(z, z') > 0$ and all measurable functions $f : \mathbb{Z}^2 \rightarrow \mathbb{R}_+$,*

$$(6.16) \quad \hat{R}((z, z'); f) = G(\phi(z, z'); \tilde{f}) \quad \text{with} \quad \tilde{f}(y) = f(\Psi_y(z), \Psi_y(z')).$$

Let us complete the proof of Theorem 9 before proving this lemma. By Lemma 20 and Lemma 19, it remains to check that (6.12) and (6.14) hold for all $(z, z') \in \mathbb{Z}^2$. Observe that by definition of \hat{R} , we have for all $n \in \mathbb{Z}_+^*$, $(z, z') \in \mathbb{Z}^2$ and measurable function $f : \mathbb{Z}^2 \rightarrow \mathbb{R}$,

$$\hat{R}^n((z, z'); f) = J_n((z, z'); f).$$

So, condition [\(AG-7\)\(iv\)](#) can be equivalently stated as: there exist $m \in \mathbb{Z}_+^*$ and $D > 0$ such that

$$\sup_{(z, z') \in \mathbb{Z}^2} \left(\hat{R}^m((z, z'); W) - W(z, z') \right) < \infty$$

and

$$\sup_{(z, z') \in \mathbb{Z}^2} \left(\hat{R}((z, z'); W) - DW(z, z') \right) < \infty.$$

Thus we can find $(\beta, D_2) \in \mathbb{R} \times (0, \infty)$ such that [\(6.14\)](#) and [\(6.15\)](#) hold for all $(z, z') \in \mathbb{Z}^2$.

Now, let $(z, z') \in \mathbb{Z}^2$. When $z = z'$, then for all $n \in \mathbb{Z}_+^*$ and $y_{1:n} \in \mathbb{Y}^n$, $d_{\mathbb{Z}}(\Psi\langle y_{1:n} \rangle(z), \Psi\langle y_{1:n} \rangle(z')) = 0$, implying that [\(6.12\)](#) and [\(6.13\)](#) hold with any nonnegative D_1 and ϱ . For $z \neq z'$, let D_1 and ϱ be defined respectively by

$$D_1 = \bigvee_{k=1}^{\ell-1} \sup_{\substack{(z, z', y_{1:k}) \in \mathbb{Z}^2 \times \mathbb{Y} \\ z \neq z'}} \frac{d_{\mathbb{Z}}(\Psi\langle y_{1:k} \rangle(z), \Psi\langle y_{1:k} \rangle(z'))}{d_{\mathbb{Z}}(z, z')},$$

and

$$\varrho = \sup_{\substack{(z, z', y_{1:\ell}) \in \mathbb{Z}^2 \times \mathbb{Y} \\ z \neq z'}} \frac{d_{\mathbb{Z}}(\Psi\langle y_{1:\ell} \rangle(z), \Psi\langle y_{1:\ell} \rangle(z'))}{d_{\mathbb{Z}}(z, z')}.$$

By [\(AG-6\)](#), we have $D_1 \in (0, \infty)$ and $\varrho \in (0, 1)$. Therefore, Conditions [\(6.12\)](#) and [\(6.13\)](#) hold for all $(z, z') \in \mathbb{Z}^2$ with the above D_1 and ϱ . \square

We conclude this section with the postponed

PROOF OF LEMMA 20. Let $(z, z') \in \mathbb{Z}^2$. We define $\bar{H}((z, z'); \cdot)$ as the distribution of (Y, Y', U) drawn as follows. We first draw a random variable \bar{Y} taking values in \mathbb{Y} with density $g(\phi(z, z'); \cdot)$ with respect to ν . Then we define (Y, Y', U) by separating the two cases, $\alpha(z, z') = 1$ and $\alpha(z, z') < 1$.

- Suppose that $\alpha(z, z') = 1$. Then from [\(AG-7\)\(i\)](#), we have

$$H(z; \cdot) = H(z'; \cdot) = G(\phi(z, z'); \cdot).$$

In this case, we set $(Y, Y', U) = (\bar{Y}, \bar{Y}, 1)$.

- Suppose now that $\alpha(z, z') < 1$. Then, using [\(3.26\)](#), the functions

$$(1 - \alpha(z, z'))^{-1} [g(\Pi_p(z); \cdot) - \alpha(z, z')g(\phi(z, z'); \cdot)]$$

and

$$(1 - \alpha(z, z'))^{-1} [g(\Pi_p(z'); \cdot) - \alpha(z, z')g(\phi(z, z'); \cdot)],$$

are probability density functions with respect to ν and we let Λ and Λ' be two independent random variables taking values in Y drawn with these two density functions, respectively. In this case we draw U independently according to a Bernoulli variable with mean $\alpha(z, z')$ and set

$$(Y, Y') = \begin{cases} (\bar{Y}, \bar{Y}) & \text{if } U = 1, \\ (\Lambda, \Lambda') & \text{if } U = 0. \end{cases}$$

One can easily check that the so defined kernel \bar{H} satisfies (6.6) and (6.7). Moreover, for all $(z, z') \in \mathsf{Z}^2$,

$$\bar{H}((z, z'); \mathsf{Y}^2 \times \{1\}) = \mathbb{P}(U = 1) = \alpha(z, z'),$$

which is compatible with (6.8). The kernel \hat{R} is defined by setting $\hat{R}((z, z'); \cdot)$ as the conditional distribution of $(Z, Z') = (\Psi_Y(z), \Psi_Y(z'))$ given that $U = 1$. To complete the proof of Lemma 20, observe that for any measurable $f : \mathsf{Z}^2 \rightarrow \mathbb{R}_+$, we have, for all $(z, z') \in \mathsf{Z}^2$ such that $\alpha(z, z') > 0$,

$$\begin{aligned} \hat{R}((z, z'); f) &= \mathbb{E} [f(\Psi_Y(z), \Psi_Y(z')) \mid U = 1] \\ &= \mathbb{E} [f(\Psi_{\bar{Y}}(z), \Psi_{\bar{Y}}(z'))] \\ &= G(\phi(z, z'); \tilde{f}), \end{aligned}$$

where $\tilde{f}(y) = f(\Psi_y(z), \Psi_y(z'))$ for all $y \in \mathsf{Y}$. □

6.3. *Proof of Lemma 11.* If $a_i \geq 0$, $b_i \geq 0$, for all $i \in \{1, \dots, p\}$, and $\sum_{i=1}^p (a_i + b_i) < 1$, it is trivial that $|\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T) < 1$. Otherwise, there would exist some λ such that $|\lambda|^{-1} \leq 1$ and

$$\left| \sum_{i=1}^p (a_i + b_i) \lambda^{-i} \right| \leq \sum_{i=1}^p (a_i + b_i) |\lambda|^{-i} < 1.$$

This would imply $P(\lambda) = \lambda^{2p-1} (1 - \sum_{k=1}^p (a_k + b_k) \lambda^{-k}) \neq 0$. Suppose on the other hand that $\lambda^* = |\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T) < 1$. From (4.1), we can write $P(\lambda)$ as a product of λ^{p-1} and $Q(\lambda) = \lambda^p - \sum_{k=1}^p (a_k + b_k) \lambda^{p-k}$. Note that $\lambda^* = 0$ if and only if $\sum_{i=1}^p (a_i + b_i) = 0$. Then we can suppose that $0 < \lambda^* < 1$, then there exists $j \in \{1, \dots, p\}$ such that $a_j + b_j > 0$. Furthermore, note that for all $i \in \{1, \dots, p\}$, $(a_i + b_i) \lambda^{*i} \geq (a_i + b_i) \lambda^{*p}$ and this inequality is strict when $i = j$. Thus from Perron-Frobenius theorem for nonnegative matrix, $Q(\lambda^*) = 0$, and we have

$$0 < \lambda^{*p} \left(1 - \sum_{i=1}^p (a_i + b_i) \right).$$

This completes the proof since $\lambda^* > 0$.

6.4. *Proof of Lemma 13.* If $\vartheta \leq 0$, then (4.11) holds since $(1 + Y)^\vartheta \leq 1$. Note that for $s \in (0, 1]$, we have

$$\mathbb{E}[(1 + Y)^s] \leq (\mathbb{E}[1 + Y])^s \leq (1 + \zeta)^s \leq 1 + \zeta^s.$$

Thus, (4.11) holds when $\vartheta \in (0, 1]$. For $\vartheta \in (1, 3/2]$, then there exists $s_1 \in (0, 1/2]$ such that $\vartheta = 1 + s_1$. By Cauchy-Schwartz inequality, we have

$$\begin{aligned} \mathbb{E}[(1 + Y)^\vartheta] &= \mathbb{E}[(1 + Y)(1 + Y)^{s_1}] \leq (\mathbb{E}[(1 + Y)^2])^{1/2} (\mathbb{E}[(1 + Y)^{2s_1}])^{1/2} \\ &\leq (1 + 3\zeta + \zeta^2)^{1/2} (1 + \zeta)^{s_1} \leq c_1 + c_2\zeta^\vartheta, \end{aligned}$$

for some constants $c_1, c_2 > 0$. If $\vartheta \in (3/2, 2]$, then there exists $s_2 \in (0, 1/2]$ such that $\vartheta = 3/2 + s_2$ and

$$\begin{aligned} \mathbb{E}[(1 + Y)^\vartheta] &= \mathbb{E}[(1 + Y)^{3/2}(1 + Y)^{s_2}] \leq (\mathbb{E}[(1 + Y)^3])^{1/2} (\mathbb{E}[(1 + Y)^{2s_2}])^{1/2} \\ &\leq (1 + 7\zeta + 6\zeta^2 + \zeta^3)^{1/2} (1 + \zeta)^{s_2} \leq c_3 + c_4\zeta^\vartheta, \end{aligned}$$

for some constants $c_3, c_4 > 0$. Finally, if $\vartheta > 2$, then there exists $s_3 \in (0, 1]$ such that $\vartheta = \lfloor \vartheta \rfloor + s_3$ and we have

$$\begin{aligned} \mathbb{E}[(1 + Y)^\vartheta] &= \mathbb{E}[(1 + Y)^{\lfloor \vartheta \rfloor} (1 + Y)^{s_3}] \\ &\leq \left(\mathbb{E}[(1 + Y)^{2\lfloor \vartheta \rfloor}] \right)^{1/2} (\mathbb{E}[(1 + Y)^{2s_3}])^{1/2} \\ &\leq \left(1 + 2^{\lfloor \vartheta \rfloor} \left(\mathbb{E}[Y^{2\lfloor \vartheta \rfloor}] \right)^{1/2} \right) (c_5 + c_6\zeta^{s_3}), \end{aligned}$$

for some constants $c_5, c_6 > 0$. The proof then follows by observing that

$$\lim_{\zeta \rightarrow \infty} \frac{(\mathbb{E}[Y^{2\lfloor \vartheta \rfloor}])^{1/2}}{\zeta^{\lfloor \vartheta \rfloor}} = 1.$$

LEMMA 21. *Let $\{U_k : k \in \mathbb{Z}_+\}$ be a stationary sequence of real-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that $\mathbb{E}(\ln^+ |U_0|) < \infty$. Then, for all $\eta \in (0, 1)$,*

$$\lim_{k \rightarrow \infty} \eta^k U_k = 0, \quad \mathbb{P}\text{-a.s.}$$

PROOF. See (Douc, Doukhan and Moulines, 2013, Lemma 34). □

References.

- BHASKARAN, K., GASPARRINI, A., HAJAT, S., SMEETH, L. and ARMSTRONG, B. (2013). Time series regression studies in environmental epidemiology. *International journal of epidemiology* dyt092.
- BOLLERSLEV, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *J. Econometrics* **31** 307–327.
- BOLLERSLEV, T. (2008). Glossary to ARCH (GARCH) Technical Report, CREATES Research Paper 2008–49.
- BOUGEROL, P. and PICARD, N. (1992). Stationarity of Garch processes and of some nonnegative time series. *J. Econometrics* **52** 115 - 127.
- CHRISTOU, V. and FOKIANOS, K. (2014). Quasi-Likelihood Inference for Negative Binomial Time Series Models. *Journal of Time Series Analysis* **35** 55–78.
- CHRISTOU, V. and FOKIANOS, K. (2015). On count time series prediction. *Journal of Statistical Computation and Simulation* **85** 357–373.
- COMTE, F. and LIEBERMAN, O. (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* **84** 61–84.
- COX, D. (1981). Statistical analysis of time-series: some recent developments. *Scand. J. Statist.* **8** 93–115.
- DAVIS, R. A., DUNSMUIR, W. and WANG, Y. (1999). Modeling time series of count data. *STATISTICS TEXTBOOKS AND MONOGRAPHS* **158** 63–114.
- DAVIS, R., DUNSMUIR, W. and STRETT, S. (2003). Observation-driven models for Poisson counts. *Biometrika* **90** 777–790.
- DAVIS, R. A. and LIU, H. (2012). Theory and Inference for a Class of Observation-Driven Models with Application to Time Series of Counts. *Preprint, arXiv:1204.3915*.
- DOUC, R., DOUKHAN, P. and MOULINES, E. (2013). Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stoch. Proc. App.* **123** 2620–2647.
- DOUC, R., ROUEFF, F. and SIM, T. (2015). Handy sufficient conditions for the convergence of the maximum likelihood estimator in observation-driven models. *Lithuanian Mathematical Journal* **55** 367–392.
- DOUC, R., ROUEFF, F. and SIM, T. (2016). The maximizing set of the asymptotic normalized log-likelihood for partially observed Markov chains. *Annals of Applied Probability* **26** 2357–2383.
- DOUKHAN, P., FOKIANOS, K. and TJØSTHEIM, D. (2012). On weak dependence conditions for Poisson autoregressions. *Statist. Probab. Lett.* **82** 942–948.
- FERLAND, R., LATOUR, A. and ORAICHI, D. (2006). Integer-valued GARCH process. *J. Time Ser. Anal.* **27** 923–942. [MR2328548](#)
- FOKIANOS, K., NEUMANN, M. H. et al. (2013). A goodness-of-fit test for Poisson count processes. *Electronic Journal of Statistics* **7** 793–819.
- FOKIANOS, K., RAHBEK, A. and TJØSTHEIM, D. (2009). Poisson autoregression. *J. Am. Statist. Assoc.* **104** 1430–1439. With electronic supplementary materials available online. [MR2596998](#) (2011d:62256)
- FOKIANOS, K. and TJØSTHEIM, D. (2011). Log-linear Poisson autoregression. *J. of Multivariate Analysis* **102** 563–578.
- FOKIANOS, K. and TJØSTHEIM, D. (2012). Nonlinear poisson autoregression. *Annals of the Institute of Statistical Mathematics* **64** 1205–1225.
- FRANCQ, C. and ZAKOIAN, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10** 605–637.
- FRANCQ, C. and ZAKOIAN, J.-M. (2009). A tour in the asymptotic theory of GARCH

- estimation. *Handbook of Financial Time Series* 85–111.
- FRANCO, C. and ZAKOIAN, J.-M. (2011). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.
- FRANKE, J. (2010). *Weak dependence of functional INGARCH processes*. Technische Universität Kaiserslautern, Fachbereich Mathematik.
- HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and its Applications*. Academic Press, New York, London.
- HEINEN, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. *Available at SSRN 1117187*.
- HENDERSON, S. G., MATTESON, D. S. and WOODARD, D. B. (2011). Stationarity of generalized autoregressive moving average models. *Electronic Journal of Statistics* **5** 800–828.
- IVES, A., DENNIS, B., COTTINGHAM, K. and CARPENTER, S. (2003). Estimating community stability and ecological interactions from time-series data. *Ecological monographs* **73** 301–330.
- LEROUX, B. G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.* **40** 127–143.
- LIESENFELD, R. and RICHARD, J.-F. (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of empirical finance* **10** 505–531.
- LINDNER, A. M. (2009). Stationarity, mixing, distributional properties and moments of GARCH (p, q)–processes. In *Handbook of financial time series* 43–69. Springer.
- MEYN, S. P. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press, London.
- NEUMANN, M. H. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **17** 1268–1284.
- PINDYCK, R. S. and RUBINFELD, D. L. (1998). *Econometric models and economic forecasts* **4**. Irwin/McGraw-Hill Boston.
- RYDBERG, T. H. and SHEPHARD, N. (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics* **1** 2–25.
- STREETT, S. (2000). Some observation driven models for time series of counts PhD thesis, Colorado State University, Department of Statistics.
- TJØSTHEIM, D. (2015). *Count Time Series with Observation-Driven Autoregressive Parameter Dynamics*. Handbook of Discrete-Valued Time Series. London: Chapman & Hall.
- WANG, C., LIU, H., YAO, J.-F., DAVIS, R. A. and LI, W. K. (2014). Self-excited threshold Poisson autoregression. *Journal of the American Statistical Association* **109** 777–787.
- ZEGER, S. L. (1988). A regression model for time series of counts. *Biometrika* **75** 621–629.

DEPARTMENT OF FOUNDATION YEAR
INSTITUTE OF TECHNOLOGY OF CAMBODIA
12156 PHNOM PENH
CAMBODIA
E-MAIL: tepmony.sim@itc.edu.kh

DÉPARTEMENT CITI
CNRS UMR 5157
TÉLÉCOM SUDPARIS
91000 ÉVRY
FRANCE
E-MAIL: randal.douc@telecom-sudparis.eu

CNRS LTCI
TÉLÉCOM PARISTECH
UNIVERSITÉ PARIS-SACLAY
75013 PARIS
FRANCE
E-MAIL: roueff@telecom-paristech.fr