HAL open science

GENERAL-ORDER OBSERVATION-DRIVEN MODELS: ERGODICITY AND CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR

Tepmony Sim, Randal Douc, François Roueff

▶ **To cite this version:**

Tepmony Sim, Randal Douc, François Roueff. GENERAL-ORDER OBSERVATION-DRIVEN MODELS: ERGODICITY AND CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR. 2016. hal-01383554v2

HAL Id: hal-01383554

https://hal.science/hal-01383554v2

Preprint submitted on 4 Apr 2019 (v2), 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: ERGODICITY AND CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR

TEPMONY SIM, RANDAL DOUC, AND FRANÇOIS ROUEFF

Abstract. The class of observation-driven models (ODMs) includes many models of non-linear time series which, in a fashion similar to, yet different from, hidden Markov models (HMMs), involve hidden variables. Interestingly, in contrast to most HMMs, ODMs enjoy likelihoods that can be computed exactly with computational complexity of the same order as the number of observations, making maximum likelihood estimation the privileged approach for statistical inference for these models. A celebrated example of general order ODMs is the GARCH(p, q) model, for which ergodicity and inference has been studied extensively. However little is known on more general models, in particular integer-valued ones, such as the log-linear Poisson GARCH or the NBIN-GARCH of order (p, q) about which most of the existing results seem restricted to the case $p = q = 1$. Here we fill this gap and derive ergodicity conditions for general ODMs. The consistency and the asymptotic normality of the maximum likelihood estimator (MLE) can then be derived using the method already developed for first order ODMs.

1. Introduction

Since they were introduced in [7], observation-driven models have been receiving renewed interest in recent years. These models are widely applied in various fields ranging from economics (see [35]), environmental study (see [1]), epidemiology and public health study (see [44, 11, 17]), finance (see [30, 36, 18, 24]) and population dynamics (see [28]). The celebrated GARCH$(1, 1)$ model introduced in [2] as well as most of the models derived from this one are typical examples of ODMs; see [3] for a list of some of them. A list of contributions on this class of models specifically dealing with discrete data includes [39, 9, 26, 17, 20, 25, 18, 27, 33, 10, 16, 12, 19, 5, 6, 8] and [14].

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 [20, 18, 21] and [43]. For a more general case where the model belongs to the class of one-parameter exponential ODMs, such as the

Date: April 4, 2019.

1991 *Mathematics Subject Classification.* Primary: 60J05, 62F12; Secondary: 62M05,62M10.

Key words and phrases. consistency, ergodicity, general-order, maximum likelihood, observation-driven models, time series of counts.

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 [10]. 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, has been obtained recently in [12], where the observed process may admit various conditional distributions. This result has later been extended and refined in [14]. 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, up to our knowledge, little is known for the GARCH(p,q)-type, i.e. larger order discrete ODMs, as highlighted as a remaining unsolved problem in [41].

Here, following others (e.g. [39, 26]), we consider a general class of ODMs that is capable to account for several lagged variables of both hidden and observation processes. We develop a theory for the class of general-order ODMs parallel to the GARCH(p,q) family. Our main contribution is to provide a complete set of conditions for general order ODMs for proving the ergodicity of the process and the consistency of the MLE, under the assumption of well-specified models. In principle, the general order model can be treated by embedding it into a first-order one and by applying the results obtained e.g. in [12, 14] to the embedded model. Yet the particular form of the embedded model does not fit the usual assumptions tailored for standard first-order ODMs. Here we derive efficient conditions taking advantage of the asymptotic behavior of iterated versions of the kernels involved. Incidentally, this also allows us to improve known conditions for some first-order models, as explained in Remark 2-(4) below. To demonstrate the efficiency of our approach, we apply our results to two specific integer valued ODMs, namely, the log-linear Poisson GARCH(p,q) model and the NBIN-GARCH(p,q). Numerical experiments involving these models can be found in [37, Section 5.5] for a set of earthquake data from Earthquake Hazards Program [40].

The paper is structured as follows. Definitions used throughout the paper are introduced in Section 2, where we also state our results on two specific examples. In Section 3, we present our main results on the ergodicity and consistency of the MLE for general order ODMs. Finally, Section 4 contains the postponed proofs and we gather some independent useful lemmas in Section 5.

2. Definitions and examples

2.1. **Observation-driven model of order** (p,q)**.** Throughout the paper we use the notation $u_{\ell:m} := (u_\ell, \dots, u_m)$ for $\ell \le m$, with the convention that $u_{\ell:m}$ is the empty sequence if $\ell > m$, so that, for instance $(x_{0:(-1)}, y) = y$. The observation-driven time series model can formally be defined as follows.

Definition 1 (General order ODM and LODM)**.** Let $(\mathsf{X}, \mathcal{X})$, $(\mathsf{Y}, \mathcal{Y})$ and $(\mathrm{U}, \mathcal{U})$ be measurable spaces, respectively called the *latent space*, the *observation space* and the *reduced observation space*. Let (Θ, Δ) be a compact metric space, called the *parameter space*. Let Υ be a measurable function from $(\mathsf{Y}, \mathcal{Y})$ to $(\mathrm{U}, \mathcal{U})$. Let $\left\{(x_{1:p}, u_{1:q}) \mapsto \tilde{\psi}^\theta_{u_{1:q}}(x_{1:p}) : \theta \in \Theta\right\}$ be a family of measurable functions from $(\mathsf{X}^p \times \mathrm{U}^q, \mathcal{X}^{\otimes p} \otimes \mathcal{U}^{\otimes q})$ to $(\mathsf{X}, \mathcal{X})$, called the *reduced link functions* and let $\left\{G^\theta : \theta \in \Theta\right\}$ be a family of probability kernels on $\mathsf{X} \times \mathcal{Y}$, called the *observation kernels*. 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 reduced link function $\tilde{\psi}^\theta$, admissible mapping Υ and observation kernel G^θ if there exists a process $\{X_k : k \geq -p+1\}$ on $(\mathsf{X}, \mathcal{X})$ such that for all $k \in \mathbb{Z}_+ = \{0,1,2,\dots\}$,

$$\begin{aligned} &Y_k \mid \mathcal{F}_k \sim G^\theta(X_k;\cdot), \\ &X_{k+1} = \tilde{\psi}^\theta_{U_{k-q+1:k}}(X_{(k-p+1):k}), \end{aligned} \tag{2.1}$$

where $U_k = \Upsilon(Y_k)$ for all $k > -q$ $\mathcal{F}_k = \sigma\left(X_{(-p+1):k}, Y_{(-q+1):(k-1)}\right)$. We further say that this model is a *linearly observation-driven model of order* (p,q) (shortened as LODM(p,q)) if moreover

(i) All $\theta \in \Theta$ can be written as $\theta = (\vartheta, \varphi)$ with $\vartheta \in \mathbb{R}^{1+p+q}$.
(ii) The latent space X is a closed subset of $\mathbb{R}$, $\mathrm{U} = \mathbb{R}$ and, for all $x = x_{1:p} \in \mathsf{X}^p$, $u = u_{1:q} \in \mathbb{R}^q$, and $\theta = (\vartheta, \varphi) \in \Theta$ with $\vartheta = (\omega, a_{1:p}, b_{1:q})$,

$$\tilde{\psi}^\theta_u(x) = \omega + \sum_{i=1}^p a_i\, x_i + \sum_{i=1}^q b_i\, u_i\ . \tag{2.2}$$

Unless differently stated, we always assume that the model is dominated by a σ-finite measure ν on $(\mathsf{Y}, \mathcal{Y})$, that is, for all $\theta \in \Theta$, there exists a measurable function $g^\theta : \mathsf{X} \times \mathsf{Y} \to \mathbb{R}_+$ written as $(x,y) \mapsto g^\theta(x;y)$ such that for all $x \in \mathsf{X}$, $g^\theta(x;\cdot)$ is the density of $G^\theta(x;\cdot)$ with respect to ν. In addition, we always assume that for all $(x,y) \in \mathsf{X} \times \mathsf{Y}$ and all $\theta \in \Theta$,

$$g^\theta(x;y) > 0\ , \tag{2.3}$$

and also, to avoid a trivial degenerate case, that $\nu \circ \Upsilon^{-1}$ is non-degenerate, that is, its support contains at least two points.

Remark 1. Let us comment briefly on this definition.

(1) The standard definition of an observation driven model does not include the admissible mapping Υ and indeed, we can define the same model without Υ by replacing the second equation in (2.1) by

$$X_{k+1} = \psi^\theta_{Y_{k-q+1:k}}(X_{(k-p+1):k}),\ ,$$

where $\left\{(x_{1:p}, y_{1:q}) \mapsto \psi^\theta_{y_{1:q}}(x_{1:p}) : \theta \in \Theta\right\}$ is a family of measurable functions from $(\mathsf{X}^p \times \mathsf{Y}^q, \mathcal{X}^{\otimes p} \otimes \mathcal{Y}^{\otimes q})$ to $(\mathsf{X}, \mathcal{X})$, called the (non-reduced) *link functions*, and defined by

$$\psi^\theta_y(x) = \tilde{\psi}^\theta_{\Upsilon^{\otimes q}(y)}(x)\ , \qquad x \in \mathsf{X}^p\ ,\ y \in \mathsf{Y}^q\ . \tag{2.4}$$

However by inserting the mapping Υ, we introduce some flexibility and this is useful for describing various ODMs with the same reduced link function $\tilde{\psi}^\theta$. For instance all LODMs use the form of reduced link function in (2.2) although they may use various mappings Υ's. This is the case for the GARCH, Log linear GARCH and NBIN GARCH, see below.

(2) When $p = q = 1$, then the ODM(p,q) defined by (2.1) collapses to the (first-order) ODM considered in [12] and [14]. Note also that if $p \neq q$, setting $r := \max(p,q)$, the ODM(p,q) can be embedded in an ODM(r,r), but this requires augmenting the parameter dimension which might impact the identifiability of the model.

(3) The standard GARCH(p,q) model is a special case of LODM(p,q), in which case $\mathsf{X}=\mathbb{R}_+$, $\mathsf{Y}=\mathbb{R}$, $\Upsilon(y)=y^2$, and $G^\theta(x,\cdot)$ is a centered distribution with variance x, most commonly the standard normal distribution.

The GARCH model has been extensively studied, see, for example, [4, 22, 23, 31, 24] and the references therein. Many other examples have been derived in the class of LODMs. An important feature of the GARCH is the fact that $x\mapsto G^\theta(x;\cdot)$ maps a family of distributions parameterized by the scale parameter $\sqrt{x}$, which is often expressed by replacing the first line in (2.1) by an equation of the form $Y_k=\sqrt{X_k}\epsilon_k$ with the assumption that $\{\epsilon_n\,:\,n\in\mathbb{Z}\}$ is i.i.d. Such a simple multiplicative formula no longer holds when the observations Y_k's are integers, which seriously complicates the theoretical analysis of such models, as explained in [41]. Also it can be of interest to let θ depend on an extra parameter φ that may influence the conditional distribution $G^\theta(x;\cdot)$.

Definition 2 (Space $(\mathsf{Z},\mathcal{Z})$, Notation $\mathbb{P}^\theta_\xi$)**.** Consider an ODM as in Definition 1. Then, for all $k\ge 0$, the conditional distribution of (Y_k,X_{k+1}) given $\mathcal{F}_k$ only depends on

$$Z_k=\left(X_{(k-p+1):k},U_{(k-q+1):(k-1)}\right)\in\mathsf{Z}\,, \tag{2.5}$$

where we defined

$$\mathsf{Z}=\mathsf{X}^p\times\mathrm{U}^{q-1}\quad\text{endowed with the }\sigma\text{-field }\mathcal{Z}=\mathcal{X}^{\otimes p}\otimes\mathcal{U}^{\otimes(q-1)}. \tag{2.6}$$

For any probability distribution ξ on $(\mathsf{Z},\mathcal{Z})$, we denote by $\mathbb{P}^\theta_\xi$ the distribution of $\{X_k,Y_{k'}\,:\,k>-p,\ k'>-q\}$ satisfying (2.1) with $Z_0\sim\xi$, with the usual notation $\mathbb{P}^\theta_z$ in the case where ξ is a Dirac mass at $z\in\mathsf{Z}$.

The inference about the model parameter is classically performed by relying on the likelihood of the observations $(Y_0,\dots,Y_n)$ given Z_0. For all $z\in\mathsf{Z}$, the corresponding conditional density function $p^\theta(y_{0:n}|z)$ with respect to $\nu^{\otimes n}$ under parameter $\theta\in\Theta$ is given by

$$y_{0:n}\mapsto\prod_{k=0}^{n}g^\theta\left(x_k;y_k\right), \tag{2.7}$$

where the sequence $x_{0:n}$ is defined through the initial conditions and recursion equations

$$\begin{cases} x_k=\Pi_{p+k}\left(z\right)\ , & -p<k\le 0\,,\\ u_k=\Pi_{p+q+k}\left(z\right)\ , & -q<k\le -1\,,\\ u_k=\Upsilon(y_k)\,, & 0\le k<n\,,\\ x_k=\tilde{\psi}^\theta_{u_{(k-q):(k-1)}}\left(x_{(k-p):(k-1)}\right)\ , & 1\le k\le n\,,\end{cases} \tag{2.8}$$

where, throughout the paper, for all $j\in\{1,\dots,p+q-1\}$, we denote by $\Pi_j\left(z\right)$ the j-th entry of $z\in\mathsf{Z}$. Note that x_{k+1} only depends on z and $y_{0:k}$ for all $k\ge 0$. Throughout the paper, we use the notation : for all $n\ge 1$, $y_{0:n-1}\in\mathsf{Y}^n$ and $z\in\mathsf{Z}$,

$$\psi^\theta\langle y_{0:(n-1)}\rangle(z):=x_n\ ,\ \text{with } x_n \text{ defined by (2.8)} \tag{2.9}$$

Note that for $n=-1$, $y_{0:-1}$ is the empty sequence and (2.9) is replaced by $\psi^\theta\langle\emptyset\rangle(z)=x_0=\Pi_p\left(z\right)$. Given the initial condition $Z_0=z^{(\mathrm{i})}$, the (conditional)

maximum likelihood estimator $\hat{\theta}_{z^{(\mathrm{i})},n}$ of the parameter θ is thus defined by

$$\hat{\theta}_{z^{(\mathrm{i})},n} \in \operatorname*{argmax}_{\theta\in\Theta} \mathsf{L}^{\theta}_{z^{(\mathrm{i})},n}, \tag{2.10}$$

where, for all $z^{(\mathrm{i})} \in \mathsf{Z}$,

$$\mathsf{L}^{\theta}_{z^{(\mathrm{i})},n} := \sum_{k=0}^{n} \ln g^{\theta}\left(\psi^{\theta}\langle Y_{0:(k-1)}\rangle(z^{(\mathrm{i})}); Y_k\right). \tag{2.11}$$

Note that since $\{X_n : n \in \mathbb{Z}\}$ is unobserved, to compute the conditional likelihood, we used some arbitrary initial values for $X_{(-p+1):0}$, (the first p entries of $z^{(\mathrm{i})}$), and that for convenience we also replace $Y_{(-p+1):-1}$ by some arbitrary values (the last $q-1$ entries of $z^{(\mathrm{i})}$) and index our set of observations as $Y_{0:n}$ (hence assuming $1+n$ "true" observations). The iterated function $\psi^{\theta}\langle Y_{0:k}\rangle$ can be cumbersome but is very easy to compute in practice using the recursion (2.8). Moreover, the same kind of recursion holds for its derivatives with respect to θ, allowing one to apply gradient steps to locally maximize the likelihood.

In this contribution, we investigate the convergence of $\hat{\theta}_{z^{(\mathrm{i})},n}$ as $n \to \infty$ for some (well-chosen) value of $z^{(\mathrm{i})}$ under the assumption that the model is well specified and the observations are in *the* steady state. The first problem to solve in this line of results is thus to show the following.

(A-1) For all $\theta \in \Theta$, there exists a unique stationary solution satisfying (2.1).

This ergodic property is the cornerstone for making statistical inference theory work and we provide simple general conditions in Section 3.2. We now introduce the notation that will allow us to refer to the stationary distribution of the model throughout the paper.

Definition 3. If (A-1) holds, then

a) $\mathbb{P}^{\theta}$ denotes the distribution on $((\mathsf{X}\times\mathsf{Y})^{\mathbb{Z}}, (\mathcal{X}\times\mathcal{Y})^{\otimes\mathbb{Z}})$ of the stationary solution of (2.1) extended on $k \in \mathbb{Z}$, with $\mathcal{F}_k = \sigma(X_{-\infty:k}, Y_{-\infty:(k-1)})$;
b) $\tilde{\mathbb{P}}^{\theta}$ denotes the projection of $\mathbb{P}^{\theta}$ on the component $\mathsf{Y}^{\mathbb{Z}}$.

We also use the symbols $\mathbb{E}^{\theta}$ and $\tilde{\mathbb{E}}^{\theta}$ to denote the expectations corresponding to $\mathbb{P}^{\theta}$ and $\tilde{\mathbb{P}}^{\theta}$, respectively. We further denote by $\pi^{\theta}_{\mathsf{X}}$ and $\pi^{\theta}_{\mathsf{Y}}$ the marginal distributions of X_0 and Y_0 under $\mathbb{P}^{\theta}$, on $(\mathsf{X},\mathcal{X})$ and $(\mathsf{Y},\mathcal{Y})$, 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 [29] as an alternative to the classical identifiability condition. Namely, we say that $\hat{\theta}_{z^{(\mathrm{i})},n}$ is *equivalence-class consistent* at the true parameter $\theta_\star$ if

$$\lim_{n\to\infty} \Delta(\hat{\theta}_{z^{(\mathrm{i})},n}, [\theta_\star]) = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \tag{2.12}$$

Identifiabity can then be treated as a separate problem which consists in determining all the parameters $\theta_\star$ for which $[\theta_\star]$ is reduced to the singleton $\{\theta_\star\}$, so that equivalence-class consistency becomes the usual consistency at and only at these parameters. The identifibility problem is treated in [15] for general order ODMs. In the case of an LODM, for $\theta_\star = (\vartheta^*, \varphi^*)$ with $\vartheta^* = (\omega^*, a^\star_{1:p}, b^\star_{1:q})$, the following conditions for having $[\theta_\star] = \{\theta_\star\}$ are derived and will also be useful hereafter.

(I-1) For all $\theta = (\vartheta, \varphi) \in \Theta$ with $\vartheta = (\omega, a_{1:p}, b_{1:q})$, we have $a_{1:p} \in \mathcal{S}_p$.

(I-2) The polynomials $\mathrm{P}_p(\cdot; a^\star_{1:p})$ and $\mathrm{Q}_q(\cdot; b^\star_{1:q})$ have no common complex roots.

In these assumptions, we used the following definitions:

$$\mathcal{S}_p = \left\{ c_{1:p} \in \mathbb{R}^p : \forall z \in \mathbb{C}, |z| \le 1 \text{ implies } 1 - \sum_{k=1}^{p} c_k z^k \neq 0 \right\} , \tag{2.13}$$

$$\mathrm{P}_p(z; a_{1:p}) = z^p - \sum_{k=1}^{p} a_k z^{p-k} \quad \text{and} \quad \mathrm{Q}_q(z; b_{1:q}) = \sum_{k=0}^{q-1} b_{k+1}\, z^{q-1-k} .$$

Condition (I-1) is often referred to as the *invertibility* condition, see [38].

As a byproduct of the proof of (A-1), one usually obtains a function $V_{\mathsf{X}} : \mathsf{X} \to \mathbb{R}_+$ of interest, common to all $\theta \in \Theta$, such that the following property holds on the stationary distribution (see Section 3.2).

(A-2) For all $\theta \in \Theta$, $\pi^\theta_{\mathsf{X}}(V_{\mathsf{X}}) < \infty$.

It is here stated as an assumption for convenience. Note also that, in the following, for $V : \mathsf{X} \to \mathbb{R}_+$ and $f : \mathsf{X} \to \mathbb{R}_+$, we denote the V-norm of f by

$$|f|_V = \sup \left\{ \frac{|f(x)|}{V(x)} : x \in \mathsf{X} \right\} ,$$

with the convention $0/0 = 0$ and we write $f \lesssim V$ if $|f|_V < \infty$. With this notation, under (A-2), for any $f : \mathsf{X} \to \mathbb{R}$ such that $f \lesssim V_{\mathsf{X}}$, $\pi^\theta_{\mathsf{X}}(|f|) < \infty$ holds, and similarly, since $\pi^\theta_{\mathsf{Y}} = \pi^\theta_{\mathsf{X}} G^\theta$ as a consequence of (2.1), for any $f : \mathsf{Y} \to \mathbb{R}$ such that $G^\theta(|f|) \lesssim V_{\mathsf{X}}$, we have $\pi^\theta_{\mathsf{Y}}(|f|) < \infty$.

2.2. **Two examples.** To motivate our general results we first explicit two models of interest, namely, the log-linear Poisson GARCH(p, q) and the NBIN-GARCH(p, q) 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, q) and NBIN-GARCH(p, q) models have not been derived so far. Formally, our first example is the following.

Example 1. The Log-linear Poisson GARCH(p, q) Model is an LODM(p, q) parameterized by $\theta = (\omega, a_{1:p}, b_{1:q}) \in \Theta \subset \mathbb{R}^{p+q+1}$ with observations space $\mathsf{Y} = \mathbb{Z}_+$ and hidden variables space $\mathsf{X} = \mathbb{R}$, and with admissible mapping $\Upsilon(y) = \ln(1 + y)$, and $G(x, \cdot)$ is the Poisson distribution with mean e^x, that is, ν is the counting measure on $\mathbb{Z}_+$ and $g^\theta(x; y) = \mathrm{e}^{x\,y - \mathrm{e}^x}/(y!)$.

Our next example is the NBIN-GARCH(p, q).

Example 2. The NBIN-GARCH(p, q) model is an LODM(p, q) parameterized by $\theta = (\omega, a_{1:p}, b_{1:q}, r) \in \Theta \subset \mathbb{R}^*_+ \times \mathbb{R}^{p+q}_+ \times \mathbb{R}^*_+$ with observations space $\mathsf{Y} = \mathbb{Z}_+$ and hidden variables space $\mathsf{X} = \mathbb{R}_+$, and with admissible mapping $\Upsilon(y) = y$, and $G(x, \cdot)$ is the the negative binomial distribution with shape parameter $r > 0$ and mean $r\,x$, that is, ν is the counting measure on $\mathbb{Z}_+$ and

$$g^\theta(x; y) = \frac{\Gamma(r + y)}{y\,!\,\Gamma(r)} \left(\frac{1}{1 + x} \right)^r \left(\frac{x}{1 + x} \right)^y . \tag{2.14}$$

Our condition for showing the ergodicity of general order log-linear Poisson GARCH models requires the following definition. For all $x = x_{(1-(p\vee q)):0} \in \mathbb{R}^{p\vee q}$,

$m \in \mathbb{Z}_+$, and $w = w_{(-q+1):m} \in \{0,1\}^{q+m}$, define $\widehat{\psi}^\theta\langle w\rangle(x)$ as x_{m+1} obtained by the recursion

$$x_k = \sum_{j=1}^{p} a_j \; x_{k-j} + \sum_{j=1}^{q} b_j \; w_{k-j} \; x_{k-j} \; , \qquad 1 \le k \le m+1 \tag{2.15}$$

We can now state our result on general order log-linear Poisson GARCH models.

Theorem 4. *Consider the log-linear Poisson GARCH(p,q) model, which satisfies Eq. (2.1) under the setting of Example 1. Suppose that, for all $\theta \in \Theta$, we have*

$$\lim_{m\to\infty} \max\left\{\left|\widehat{\psi}^\theta\langle w\rangle(x)\right| : w \in \{0,1\}^{q+m}\right\} = 0 \quad \textit{for all } x \in \mathbb{R}^{p\vee q} \; . \tag{2.16}$$

Then we have the following.

(i) For all $\theta \in \Theta$, there exists a unique stationary solution $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$ to (2.1), that is, (A-1) holds. Moreover, for any $\tau > 0$, (A-2) holds with $V_{\mathsf{X}} : \mathbb{R} \to \mathbb{R}_+$ defined by

$$V_{\mathsf{X}}(x) = \mathrm{e}^{\tau|x|}, \qquad x \in \mathbb{R}. \tag{2.17}$$

(ii) Let $\theta_\star \in \Theta$. For any $x_1^{(i)} \in \mathbb{R}$ and $y_1^{(i)} \in \mathbb{Z}_+$, setting $z^{(i)} = (x_1^{(i)}, \dots, x_1^{(i)}, \Upsilon(y_1^{(i)}), \cdots, \Upsilon(y_1^{(i)})) \in \mathbb{R}^p \times \mathbb{R}^{q-1}$, the MLE $\hat{\theta}_{z^{(i)},n}$ as defined by (2.10) is equivalence-class consistent.

(iii) If the true parameter $\theta_\star = (\omega^\star, a^\star_{1:p}, b^\star_{1:q})$ moreover satisfies (I-2), then the MLE $\hat{\theta}_{z^{(i)},n}$ is consistent.

The proof is postponed to Section 4.5.

Remark 2. Let us provide some insights about Condition (2.16).

(1) Using the two possible constant sequences w, $w_k = 0$ for all k or $w_k = 1$ for all k in (2.15), we easily see that (2.16) implies

$$a_{1:p} \in \mathcal{S}_p \quad \text{and} \quad a_{1:(p\vee q)} + b_{1:(p\vee q)} \in \mathcal{S}_{p\vee q} \; ,$$

where we used the usual convention $a_k = 0$ for $p < k \le q$ and $b_k = 0$ for $q < k \le p$.

(2) A sufficient condition to have (2.16) is

$$\sup\left\{\left|\widehat{\psi}^\theta\langle w\rangle(x)\right| : w \in \{0,1\}^q, \, x \in [-1,1]^{p\vee q}\right\} < 1 \; . \tag{2.18}$$

Indeed, defining ρ as the left-hand side of the previous display, we clearly have, for all $m > p \vee q$, $w \in \{0,1\}^{q+m}$ and $x \in \mathbb{R}^{p\vee q}$,

$$\left|\widehat{\psi}^\theta\langle w\rangle(x)\right| \le \rho \max\left\{\left|\widehat{\psi}^\theta\langle w'\rangle(x)\right| : w' \in \{0,1\}^{q+m-j}, 1 \le j \le p\vee q\right\} \; .$$

(3) The first iteration (2.15) implies, for all $w \in \{0,1\}^q$ and $x \in [-1,1]^{p\vee q}$,

$$|\widehat{\psi}^\theta\langle w\rangle(x)| \le \left[\sum_{k=1}^{p\vee q} (|a_k| \vee |a_k + b_k|)\right] \; .$$

Hence a sufficient condition to have (2.18) (and thus (2.16)) is

$$\sum_{k=1}^{p\vee q} (|a_k| \vee |a_k + b_k|) < 1 \; . \tag{2.19}$$

(4) When $p = q = 1$, by Points (1) and (3) above, Condition (2.16) is equivalent to have $|a_1| < 1$ and $|a_1 + b_1| < 1$. This condition is weaker than the one derived in [12] where $|b_1| < 1$ is also imposed.

We now state our result for general order NBIN-GARCH model.

Theorem 5. *Consider the NBIN-GARCH(p, q) model, which satisfies Eq. (2.1) under the setting of Example 2. Suppose that, for all $\theta = (\omega, a_{1:p}, b_{1:q}, r) \in \Theta$, we have*

$$\sum_{k=1}^{p} a_k + r \sum_{k=1}^{q} b_k < 1. \tag{2.20}$$

Then the following assertions hold.

(i) For all $\theta \in \Theta$, there exists a unique stationary solution to (2.1), that is, (A-1) holds. Moreover, (A-2) holds with $V_{\mathsf{X}}(x) = x$ for all $x \in \mathbb{R}_+$.

(ii) Let $\theta_\star \in \Theta$. For any $x_1^{(i)} \in \mathbb{R}$ and $y_1^{(i)} \in \mathbb{Z}_+$, setting $z^{(i)} = (x_1^{(i)}, \dots, x_1^{(i)}, y_1^{(i)}, \cdots, y_1^{(i)}) \in \mathbb{R}_+^p \times \mathbb{Z}_+^{q-1}$, the MLE $\hat{\theta}_{z^{(i)},n}$ as defined by (2.10) is equivalence-class consistent.

(iii) If the true parameter $\theta_\star = (\omega^\star, a_{1:p}^\star, b_{1:q}^\star, r^\star)$ moreover satisfies Condition (I-2), then the MLE $\hat{\theta}_{z^{(i)},n}$ is consistent.

The proof is postponed to Section 4.6.

Remark 3. Clearly, under the setting of Example 2, if Eq. (2.1) has a stationary solution such that $\mu = \int x \, \pi_{\mathsf{X}}(\mathrm{d}x) < \infty$, taking the expectation on both sides of the second equation in (2.1) and using that $\int y \, \pi_{\mathsf{Y}}(\mathrm{d}y) = r \int x \, \pi_{\mathsf{X}}(\mathrm{d}x)$, then (2.20) must hold, in which case we get

$$\mu = \left(1 - \sum_{k=1}^{p} a_k + r \sum_{k=1}^{q} b_k\right)^{-1}.$$

Hence (2.20) is in fact necessary and sufficient to get a stationary solution admitting a finite first moment, as was already observed in [45, Theorem 1] although the ergodicity is not proven in this reference. However, we believe that, similarly to the classical GARCH(p, q) processes, we can find stationary solutions to Eq. (2.1) in the case where (2.20) does not hold. This is left for future work.

Remark 4. The consistency of the MLE in Theorems 4 and 5 paves the way for investigating its asymptotic normality. This is done for these two examples in [37, Proposition 5.4.7(iii) and Proposition 5.4.15].

3. General Results

3.1. **Preliminaries.** In the well-specified setting, a general result on the consistency of the MLE for a class of first-order ODMs has been obtained in [12]. Let us briefly describe the approach used to establish the convergence of the MLE $\hat{\theta}_{z^{(i)},n}$ in this reference and in the present contribution for higher order ODMs. Let $\theta_\star \in \Theta$ denote the true parameter. The consistency of the MLE is obtained through the following steps.

Step 1 Find sufficient conditions for the ergodic property (A-1) of the model. Then the convergence of the MLE to $\theta_\star$ is studied under $\widetilde{\mathbb{P}}^{\theta_\star}$ as defined in Definition 3.

Step 2 Establish that, as the number of observations $n \to \infty$, the normalized log-likelihood $\mathsf{L}^{\theta}_{z^{(\mathrm{i})},n}$ as defined in (2.11), for some well-chosen $z^{(\mathrm{i})} \in \mathsf{X}^p$, 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 $(\mathsf{Y}^{\mathbb{Z}}, \mathcal{Y}^{\otimes \mathbb{Z}})$. To define $p^{\theta}(\cdot|\cdot)$, we set, for all $y_{-\infty:0} \in \mathsf{Y}^{\mathbb{Z}_-}$ and $y \in \mathsf{Y}$, whenever the following limit is well defined,

$$p^{\theta}\left(y \,|\, y_{-\infty:0}\right) = \lim_{m \to \infty} g^{\theta}\left(\psi^{\theta}\langle y_{-m:0}\rangle(z^{(\mathrm{i})}); y\right) . \tag{3.1}$$

Step 3 By (A-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

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

Step 4 Using an additional argument (similar to that in [34]), deduce that the MLE $\hat{\theta}_{z^{(\mathrm{i})},n}$ defined by (2.10) eventually lies in any given neighborhood of the set

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

which only depends on $\theta_\star$, establishing that

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

where Δ is the metric endowing the parameter space Θ.

Step 5 Establish that $\Theta_\star$ defined in (3.2) reduces to the equivalent class $[\theta_\star]$ of Definition 3. The convergence (3.3) is then called the *equivalence class consistency* of the MLE.

Step 6 Establish that $[\theta_\star]$ reduces to the singleton $\{\theta_\star\}$. The convergence (3.3) is then called the *strong consistency* of the MLE.

In [14], we provided easy-to-check conditions on first order ODMs for obtaining **Step** 1 to **Step** 4. See [14, Theorem 2] for **Step** 1, and [14, Theorem 1] for the following steps. In [13], we proved a general result for partially observed Markov chains, which include first order ODMs, in order to get **Step** 5, see Theorem 1 in this reference. Finally **Step** 6 is often carried out using particular means adapted to the precise considered model.

We present in Section 3.2 our conditions from achieving ergodicity (**Step** 1) and, in Section 3.3, we adapt the conditions already used in [14, 13] to carry out **Step** 2 to **Step** 5 for first order model to higher order ODMs.

Using the embedding described in Section 4.1, all the steps from **Step** 1 to **Step** 5 can in principle be obtained by applying the existing results to the first order ODMs in which the original higher order model is embedded. This approach is indeed successful, up to some straightforward adaptation, for **Step** 2 to **Step** 5. Ergodicity in **Step** 1 requires a deeper analysis that constitutes the main part of this contribution. As for **Step** 6, it is treated in [15].

3.2. **Ergodicity.** In this section, we provide conditions that yield stationarity and ergodicity of the Markov chain $\{(Z_k, Y_k) : k \in \mathbb{Z}_+\}$, that is, we check (A-1) and (A-2). 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 and ψ^θ and respectively write G 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 [32]. It is not at all the same picture for integer-valued observation-driven models, where other tools need to be invoked; see [18, 12, 14] for ODMs(1,1). Here we extend these results for general order ODMs(p, q). Let us now introduce our list of assumptions. They will be further commented hereafter.

We need some metric on the space Z and assume the following.

(A-3) The σ-fields $\mathcal{X}$ and $\mathcal{U}$ are Borel ones, respectively associated to $(\mathsf{X}, \boldsymbol{\delta}_\mathsf{X})$ and $(\mathrm{U}, \boldsymbol{\delta}_\mathrm{U})$, both assumed to be complete and separable metric spaces.

Recall that any finite Y-valued sequence y, $\psi^\theta\langle y\rangle$ is defined by (2.9) with the recursion (2.8). Define, for all $n \in \mathbb{Z}_+^*$, the Lipschitz constant for $\psi^\theta\langle y\rangle$, uniform over $y \in \mathsf{Y}^n$,

$$\mathrm{Lip}_n^\theta = \sup\left\{\frac{\boldsymbol{\delta}_\mathsf{X}(\psi^\theta\langle y\rangle(z), \psi^\theta\langle y\rangle(z'))}{\boldsymbol{\delta}_\mathsf{Z}(z, z')} : (z, z', y) \in \mathsf{Z}^2 \times \mathsf{Y}^n\right\} , \tag{3.4}$$

where we set, for all $v \in \mathsf{Z}^2$,

$$\boldsymbol{\delta}_\mathsf{Z}(v) = \left(\max_{1\le k\le p} \boldsymbol{\delta}_\mathsf{X} \circ \Pi_k^{\otimes 2}(v)\right) \bigvee \left(\max_{p<k<p+q} \boldsymbol{\delta}_\mathrm{U} \circ \Pi_k^{\otimes 2}(v)\right) . \tag{3.5}$$

We use the following assumption on the link function.

(A-4) For all $\theta \in \Theta$, we have $\mathrm{Lip}_1^\theta < \infty$ and $\mathrm{Lip}_n^\theta \to 0$ as $n \to \infty$.

The following assumption is mainly related to the observation kernel G. It partly relies on the iterates of the link functions defined in (2.9), but expressed using $u_{0:(n-1)}$ instead of $y_{0:(n-1)}$. Namely, notice that x_n in (2.8) can also be defined as a measurable function of z and $u_{0:(n-1)}$, hence we can define

$$\tilde{\psi}^\theta\langle u_{0:(n-1)}\rangle(z) := x_n \ , \ \text{with } u_{0:(n-1)}, x_n \text{ defined by (2.8)} \tag{3.6}$$

so that $\psi^\theta\langle y_{0:(n-1)}\rangle(z) = \tilde{\psi}^\theta\langle \Upsilon^{\otimes n}(y_{0:(n-1)})\rangle(z)$ for all $z \in \mathsf{Z}$ and $y_{0:(n-1)} \in \mathsf{Y}^n$. The assumption on G reads as follows.

(A-5) The space $(\mathsf{X}, \boldsymbol{\delta}_\mathsf{X})$ is locally compact and if $q > 1$, so is $(\mathrm{U}, \boldsymbol{\delta}_\mathrm{U})$. For all $x \in \mathsf{X}$, there exists $\delta > 0$ such that

$$\int \sup\{g(x', y) : \boldsymbol{\delta}_\mathsf{X}(x', x) < \delta\} \ \nu(\mathrm{d}y) < \infty . \tag{3.7}$$

Moreover, one of the two following assertions hold.

(a) The kernel G is strong Feller.

(b) The kernel G is weak Feller and the function $u \mapsto \tilde{\psi}\langle u\rangle(z)$ defined in (3.6) is continuous on U for all $z \in \mathsf{Z}$.

Next, we consider a classical drift condition used for showing the existence of an invariant probability distribution.

(A-6) There exist measurable functions $V_{\mathsf{X}} : \mathsf{X} \to \mathbb{R}_+$ and $V_{\mathrm{U}} : \mathrm{U} \to \mathbb{R}_+$ such that, setting $V_{\mathsf{Y}} = V_{\mathrm{U}} \circ \Upsilon$, $GV_{\mathsf{Y}} \lesssim V_{\mathsf{X}}$, $\{V_{\mathsf{X}} \leq M\}$ is compact for any $M > 0$, and so is $\{V_{\mathsf{Y}} \leq M\}$ if $q > 1$, and

$$\lim_{n\to\infty} \lim_{M\to\infty} \sup_{z\in\mathsf{Z}} \frac{\mathbb{E}_z\left[V_{\mathsf{X}}(X_n)\right]}{M + V(z)} = 0 \ , \tag{3.8}$$

where we defined

$$V(z) = \max_{\substack{1\leq k\leq p \\ p<\ell<p+q}} \left\{ V_{\mathsf{X}}(\Pi_k\left(z\right)), \frac{V_{\mathrm{U}}(\Pi_\ell\left(z\right))}{|GV_{\mathsf{Y}}|_{V_{\mathsf{X}}}} \right\} \ . \tag{3.9}$$

The following condition is used to show the existence of a reachable point.

(A-7) One of the two following assertions hold.
  (a) There exists $y_0 \in \mathsf{Y}$ such that $\nu(\{y_0\}) > 0$.
  (b) The function $(x, y) \mapsto \tilde{\psi}_y(x)$ is continuous on $\mathsf{X}^p \times \mathsf{Y}^q$.

The last assumption is used to show the uniqueness of the invariant probability measure, through a coupling argument. It requires the following definition, used in a coupling argument. For any initial distribution ξ on $(\mathsf{Z}^2, \mathcal{Z}^{\otimes 2})$, let $\hat{\mathbb{E}}_\xi$ denote the expectation (operator) assoiciated to the distribution of $\{X_k, X'_k, Y_{k'}, Y'_{k'} : k > -p,\ k' > -q\}$ satisfying $(X_{(-p+1):0}, Y_{(-q+1):-1}, X'_{(-p+1):0}, Y'_{(-q+1):-1}) \sim \xi$ and, for all $k \in \mathbb{Z}_+$,

$$\begin{aligned} &Y_k|\mathcal{F}'_k \sim G(\phi(X_k, X'_k), \cdot) \quad \text{and} \quad Y'_k = Y_k \ , \\ &X_{k+1} = \psi_{Y_{(k-q+1):k}}(X_{(k-p+1):k}) \ , \\ &X'_{k+1} = \psi_{Y'_{(k-q+1):k}}(X'_{(k-p+1):k}) \ . \end{aligned} \tag{3.10}$$

where $\mathcal{F}'_k = \sigma\left(X_{(-p+1):k}, Y'_{(-q+1):(k-1)}, X'_{(-p+1):k}, Y_{(-q+1):(k-1)}\right)$.

(A-8) There exist measurable functions $\alpha : \mathsf{X}^2 \to [0, 1]$, $\phi : \mathsf{X}^2 \to \mathsf{X}$, a $W_{\mathsf{X}} : \mathsf{X}^2 \to [1, \infty)$ and $W_{\mathrm{U}} : \mathrm{U} \to \mathbb{R}_+$ such that, setting $W_{\mathsf{Y}} = W_{\mathrm{U}} \circ \Upsilon$, we have $(GW_{\mathsf{Y}}) \circ \phi \lesssim W_{\mathsf{X}}$ and the threee following assertions hold.
  (i) For all $(x, x') \in \mathsf{X}^2$ and $y \in \mathsf{Y}$,

$$\min\left\{g(x; y), g(x'; y)\right\} \geq \alpha(x, x') g\left(\phi(x, x'); y\right). \tag{3.11}$$

  (ii) The function W_{X} is symmetric on X^2, $W_{\mathsf{X}}(x, \cdot)$ is locally bounded for all $x \in \mathsf{X}$, and W_{U} is locally bounded on U.
  (iii) We have $1 - \alpha \leq \boldsymbol{\delta}_{\mathsf{X}} \times W_{\mathsf{X}}$ on X^2.
  And, defining, for all $v = (z, z') \in \mathsf{Z}^2$,

$$W(v) = \max\left\{ W_{\mathsf{X}} \circ \Pi_k^{\otimes 2}(v), \frac{W_{\mathrm{U}}(\Pi_\ell\left(\tilde{z}\right))}{|GW_{\mathsf{Y}} \circ \phi|_{W_{\mathsf{X}}}} : \begin{smallmatrix} 1\leq k\leq p \\ \tilde{z}\in\{z,z'\} \\ p<\ell<p+q \end{smallmatrix} \right\} \ , \tag{3.12}$$

one of the two following assertions holds.
  (iv) $\lim_{\zeta\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ \ln \sup_{v\in\mathsf{Z}^2} \frac{\hat{\mathbb{E}}_v\left[W_{\mathsf{X}}(X_n, X'_n)\right]}{W^\zeta(v)} \leq 0.$
  (v) $\lim_{n\to\infty} \lim_{M\to\infty} \sup_{v\in\mathsf{Z}^2} \frac{\hat{\mathbb{E}}_v\left[W_{\mathsf{X}}(X_n, X'_n)\right]}{M + W(z)} = 0$ and, for all $r = 1, 2, \ldots$, there exists $\tau \geq 1$ such that $\sup_{v\in\mathsf{Z}^2} \hat{\mathbb{E}}_v\left[W_{\mathsf{X}}(X_r, X'_r)\right] / W^\tau(v) < \infty$.

Remark 5. Let us comment briefly on these assumptions.

(1) By [15, Remark 2], for an LODM(p,q) (A-4) is equivalent to (I-1).
(2) If $q=1$, the terms depending on ℓ both in (3.9) and (3.12) vanish. We can take $V_{\mathrm{U}} = W_{\mathrm{U}} = 0$ without loss of generality in this case.
(3) Recall that a kernel is strong (resp. weak) Feller if it maps any bounded measurable (resp. bounded continuous) function to a bounded continuous function. By Scheffé's lemma, a sufficient condition for G to be weak Feller is to have that $x \mapsto g(x;y)$ is continuous on X for all $y \in \mathsf{Y}$. But then (3.7) gives that G is also strong Feller by dominated convergence.
(4) Note that Condition (3.7) hold when $G(x,\cdot)$ is taken among an exponential family with natural parameter continuously depending on x and valued within the open set of natural parameters, in which case G is also strong Feller. We are in this situation for both Examples 1 and 2.

We can now state the main ergodicity result.

Theorem 6. *Let $\mathbb{P}_z$ be defined as $\mathbb{P}_z^\theta$ in Definition 1. Conditions (A-3), (A-4), (A-5), (A-6), (A-7) and (A-8) imply that there exists a unique initial distribution π which makes $\mathbb{P}_\pi$ shift invariant. Moreover it satisfies $\mathbb{E}_\pi[V_{\mathsf{X}}(X_0)] < \infty$. Hence, provided that these assumptions hold at each $\theta \in \Theta$, they imply (A-1) and (A-2).*

For convenience, we postpone this proof to Section 4.3.

The following lemma provides a general way for constructing the instrumental functions α and ϕ that appear in (A-8). The proof can be easily adapted from [14, Lemma 1] and is thus omitted.

Lemma 7. *Suppose that $\mathsf{X} = \mathsf{C}^{\mathsf{S}}$ for some measurable space $(\mathsf{S}, \mathcal{S})$ and $\mathsf{C} \subseteq \mathbb{R}$. Thus for all $x \in \mathsf{X}$, we write $x = (x_s)_{s\in\mathsf{S}}$, where $x_s \in \mathsf{C}$ for all $s \in \mathsf{S}$. Suppose moreover that for all $x = (x_s)_{s\in\mathsf{S}} \in \mathsf{X}$, 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 $s \in \mathsf{S}$. This means that for all $t \in \mathsf{C}$, $y \mapsto j(t)h(t;y)$ is a density with respect to ν and there exists a probability measure μ on $(\mathsf{S}, \mathcal{S})$ such that*

$$g(x;y) = \int_{\mathsf{S}} j(x_s)h(x_s;y)\mu(\mathrm{d}s), \quad y \in \mathsf{Y}\,. \tag{3.13}$$

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

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

For all $x, x' \in \mathsf{X}^{\mathsf{S}}$, we denote $x \wedge x' := (x_s \wedge x'_s)_{s\in\mathsf{S}}$ and $x \vee x' := (x_s \vee x'_s)_{s\in\mathsf{S}}$ and we define

$$\begin{cases} \alpha(x,x') = \inf\limits_{s\in\mathsf{S}} \left\{ \dfrac{j(x_s \vee x'_s)}{j(x_s \wedge x'_s)} \right\} & \text{and} \quad \phi(x,x') = x \wedge x' \quad \text{under (H'-1)};\\ \alpha(x,x') = \inf\limits_{s\in\mathsf{S}} \left\{ \dfrac{j(x_s \wedge x'_s)}{j(x_s \vee x'_s)} \right\} & \text{and} \quad \phi(x,x') = x \vee x' \quad \text{under (H'-2)}. \end{cases}$$

Then α and ϕ defined above satisfy (A-8)(i).

3.3. **Convergence of the MLE.** Once the ergodicity of the model is established, one can derive the asymptotic behavior of the MLE, provided some regularity and moment condition holds for going through **Step** 2 to **Step** 5, as described

in Section 3.1. These steps are carried out using [14, Theorem 1] and [13, Theorem 3], written for general ODMs(1,1). The adaptation to higher order ODMs(p,q) will follow easily from the embedding of Section 4.1. We consider the following assumptions, the last of which uses V_{X} as introduced in Definition 3 under Assumptions (A-1) and (A-2).

(B-1) For all $y \in \mathsf{Y}$, the function $(\theta, x) \mapsto g^\theta(x;y)$ is continuous on $\Theta \times \mathsf{X}$.

(B-2) For all $y \in \mathsf{Y}$, the function $(\theta, z) \mapsto \psi^\theta\langle y\rangle(z)$ is continuous on $\Theta \times \mathsf{Z}$.

(B-3) There exist $x_1^{(\mathrm{i})} \in \mathsf{X}$, $y_1^{(\mathrm{i})} \in \mathsf{Y}$, a closed set $\mathsf{X}_1 \subseteq \mathsf{X}$, $C \geq 0$ and a measurable function $\bar{\phi} : \mathsf{Y} \to \mathbb{R}_+$ such that the following assertions hold setting $z^{(\mathrm{i})} = (x_1^{(\mathrm{i})}, \ldots, x_1^{(\mathrm{i})}, \Upsilon(y_1^{(\mathrm{i})}), \ldots, \Upsilon(y_1^{(\mathrm{i})})) \in \mathsf{X}^p \times \Upsilon(\mathsf{Y})^{p-1}$.

(i) For all $\theta \in \Theta$ and $(z,y) \in \mathsf{Z} \times \mathsf{Y}$, $\psi^\theta\langle y\rangle(z) \in \mathsf{X}_1$.

(ii) $\sup_{(\theta,x,y)\in\Theta\times\mathsf{X}_1\times\mathsf{Y}} g^\theta(x;y) < \infty$.

(iii) For all $y \in \mathsf{Y}$ and $\theta \in \Theta$, $\boldsymbol{\delta}_{\mathsf{X}}\left(x_1^{(\mathrm{i})}, \psi^\theta\langle y\rangle(z^{(\mathrm{i})})\right) \vee \boldsymbol{\delta}_{\mathrm{U}}(\Upsilon(y), \Upsilon(y_1^{(\mathrm{i})})) \leq \bar{\phi}(y)$.

(iv) For all $\theta \in \Theta$ and $(x, x', y) \in \mathsf{X}_1 \times \mathsf{X}_1 \times \mathsf{Y}$,

$$\left|\ln \frac{g^\theta(x;y)}{g^\theta(x';y)}\right| \leq H(\boldsymbol{\delta}_{\mathsf{X}}(x,x'))\, \mathrm{e}^{C\left(\boldsymbol{\delta}_{\mathsf{X}}(x_1^{(\mathrm{i})},x)\vee\boldsymbol{\delta}_{\mathsf{X}}(x_1^{(\mathrm{i})},x')\right)}\, \bar{\phi}(y), \tag{3.14}$$

(v) $H(u) = O(u)$ as $u \to 0$.

(vi) If $C = 0$, then, for all $\theta \in \Theta$, $G^\theta \ln^+ \bar{\phi} \lesssim V_{\mathsf{X}}$. Otherwise, for all $\theta \in \Theta$, $G^\theta \bar{\phi} \lesssim V_{\mathsf{X}}$.

Remark 6. If we consider an LODM as in Definition 1, Condition (B-2) is obvious and (B-3) (iii) reduces to impose that $\bar{\phi}(y) \geq A + B\,|\Upsilon(y)|$ for some non-negative constants A and B only depending on $x_1^{(\mathrm{i})}$ and on (the compact set) Θ.

Remark 7. In the case where the observations are discrete, one usually take ν to be the counting measure on the at most countable space Y. In this case, $g^\theta(x;y) \in [0,1]$ for all θ, x and y and Condition (B-3)(ii) trivially holds whatever X_1 is.

We have the following result, whose proof is postponed to Section 4.4.

Theorem 8. *Consider an ODM(p,q) for some $p,q \geq 1$ satisfying (A-4). Assume that (A-1), (A-2), (B-1), (B-2) and (B-3) hold. Then the MLE $\hat{\theta}_{z^{(\mathrm{i})},n}$ defined by (2.10) is equivalence-class consistent, that is, the convergence (2.12) holds for any $\theta_\star \in \Theta$.*

4. Postponed Proofs

4.1. **Embedding into an observation-driven model of order** $(1,1)$**.** A simplifying and unifying step is to embed the general order case into the order $(1,1)$ by augmenting the state space. Consider an ODM as in Definition 1. For all $y \in \mathsf{Y}$, we define $\Psi_y^\theta : \mathsf{Z} \to \mathsf{Z}$ by

$$\Psi_y^\theta : z = z_{1:(p+q-1)} \mapsto \begin{cases} \left(z_{2:p}, \psi^\theta\langle y\rangle(z), z_{(p+2):(p+q-1)}, \Upsilon(y)\right) & \text{if } q > 1 \\ \left(z_{2:p}, \psi^\theta\langle y\rangle(z)\right) & \text{if } q = 1\,. \end{cases} \tag{4.1}$$

We further denote the successive composition of $\Psi_{y_0}^\theta$, $\Psi_{y_1}^\theta$, ..., and $\Psi_{y_k}^\theta$ by

$$\Psi^\theta\langle y_{0:k}\rangle = \Psi_{y_k}^\theta \circ \Psi_{y_{k-1}}^\theta \circ \cdots \circ \Psi_{y_0}^\theta\,. \tag{4.2}$$

Note in particular that $\psi^\theta\langle y_{0:k}\rangle$ defined by (2.9) with the recursion (2.8) can be written as

$$\psi^\theta\langle y_{0:k}\rangle = \Pi_p \circ \Psi^\theta\langle y_{0:k}\rangle \ , \tag{4.3}$$

Conversely, we have, for all $k \geq 0$ and $y_{0:k} \in \mathsf{Y}^{k+1}$,

$$\Psi^\theta\langle y_{0:k}\rangle(z) = \left(\left(\psi^\theta\langle y_{0:j}\rangle(z)\right)_{k-p<j\leq k}, u_{(k-q+2):k}\right) \ , \tag{4.4}$$

where we set $u_j = \Pi_{p+q+j}(z)$ for $-q < j \leq -1$ and $u_j = \Upsilon(y_j)$ for $0 \leq j \leq k$ and use the convention $\psi^\theta\langle y_{0:j}\rangle(z) = \Pi_{p-j}(z)$ for $-p < j \leq 0$. By letting $Z_k = (X_{(k-p+1):k}, U_{(k-q+1):(k-1)})$ (see (2.5)), Model (2.1) can be rewritten as, for all $k \in \mathbb{Z}_+$,

$$\begin{aligned} &Y_k \mid \mathcal{F}_k \sim H^\theta(Z_k;\cdot), \\ &Z_{k+1} = \Psi^\theta_{Y_k}(Z_k), \end{aligned} \tag{4.5}$$

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

$$H^\theta(z;\cdot) := G^\theta(\Pi_p(z);\cdot) \ . \tag{4.6}$$

By this representation, the ODM(p,q) is thus embedded 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,q). As an ODM$(1,1)$, the bivariate process $\{(Z_k, Y_k) : k \in \mathbb{Z}_+\}$ is a Markov chain on the space $(\mathsf{Z}\times\mathsf{Y}, \mathcal{Z}\otimes\mathcal{Y})$ with transition kernel K^θ satisfying, for all $(z,y) \in \mathsf{Z}\times\mathsf{Y}$, $A \in \mathcal{Z}$ and $B \in \mathcal{Y}$,

$$K^\theta((z,y); A\times B) = \int \mathbb{1}_{A\times B}(\Psi^\theta_y(z), y')\, G^\theta(\Pi_p(z); \mathrm{d}y'). \tag{4.7}$$

Remark 8. Note that (A-1) is equivalent to saying that the transition kernel K^θ of the complete chain admits a unique invariant probability measure π^θ on $\mathsf{Z}\times\mathsf{Y}$. Moreover the resulting π^θ_{X} and π^θ_{Y} can be obtained by projecting π^θ on any of its X component and any of its Y component, respectively.

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

$$R^\theta(z; A) = \int \mathbb{1}_A(\Psi^\theta_y(z)) H^\theta(z; \mathrm{d}y) = \int \mathbb{1}_A(\Psi^\theta_y(z)) G^\theta(\Pi_p(z); \mathrm{d}y). \tag{4.8}$$

4.2. **Some additional notation.** We introduce some algebra notation that will be used hereafter. The transpose of a matrix $\mathbf{M}$ is denoted by $\mathbf{M}^T$, the identity matrix of order n by $\mathbf{I}_n$ (or simply $\mathbf{I}$), the max norm of $z \in \mathbb{R}^{p+q-1}$ by

$$|z|_\infty = \max\{\Pi_k(z) : 1 \leq k \leq p+q-1\}.$$

We further write $\boldsymbol{\varepsilon}_j$ for the j-th canonical vector in $\mathbb{R}^{p+q-1}$, $1 \leq j < p+q$, so that $\Pi_j(z) = \boldsymbol{\varepsilon}_j^T z$. For given coefficients ω, a_1, ..., a_p, b_1, ..., b_q, we define the

$(p+q-1)$-square matrix

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

and the $(p+q-1)$-dimensional vectors $\mathbf{b}$ and $\boldsymbol{\omega}$ by

$$\mathbf{b} = b_1 \boldsymbol{\varepsilon}_p + \boldsymbol{\varepsilon}_{p+q-1},$$
$$\boldsymbol{\omega} = \omega \boldsymbol{\varepsilon}_p.$$

In the case $q = 1$, we adopt the following convention: $\mathbf{A}$ reduces to its top-left p-square bloc and $\mathbf{b}$ reduces to $b_1 \boldsymbol{\varepsilon}_p$. In particular, in the case $p = q = 1$, $\mathbf{A}$, $\mathbf{b}$ and $\boldsymbol{\omega}$ reduce to $\mathbf{A} = a_1$, $\mathbf{b} = b_1$ and $\boldsymbol{\omega} = \omega$, respectively.

We further denote by $\mathbf{A}_\star$, $\mathbf{b}_\star$ and $\boldsymbol{\omega}_\star$ the corresponding values of $\mathbf{A}$, $\mathbf{b}$ and $\boldsymbol{\omega}$ at $(\omega, a_{1:p}, b_{1:q}) = (\omega^\star, a^\star_{1:p}, b^\star_{1:q})$, respectively. Recall the notation convention $\theta = (\vartheta, \varphi)$ with $\vartheta = (\omega, a_{1:p}, b_{1:q})$ in Definition 1. We use the corresponding one for $\theta_\star$: $\theta_\star = (\vartheta_\star, \varphi_\star)$ with $\vartheta_\star = (\omega^\star, a^\star_{1:p}, b^\star_{1:q})$.

4.3. **Proof of Theorem 6.** The scheme of proof of this theorem is somewhat similar to that of [14, Theorem 2] which is dedicated to the ergodicity of ODM(1,1) processes. The main difference is that we need to rely on assumptions involving *iterates* of kernels such as (E-2), (A-4) and (A-8)(iv) below, to be compared with their counterparts (A-4), (A-7) and (A-8)(iv) in [14, Theorem 2]). Using the embedding of Section 4.1, we will use the following conditions directly applying to the kernel R of this embedding.

(E-1) The space $(\mathsf{Z}, \boldsymbol{\delta}_\mathsf{Z})$ is a locally compact and complete separable metric space.

(E-2) There exists a positive integer m such that the Markov kernel R^m is weak Feller. Moreover, there exist $(\lambda, \beta) \in (0,1) \times \mathbb{R}_+$ and a measurable function $V : \mathsf{Z} \to \mathbb{R}_+$ such that $R^m V \leq \lambda V + \beta$ and $\{V \leq M\}$ is a compact set for any $M > 0$.

(E-3) The Markov kernel R admits a reachable point, that is, there exists $z_\infty \in \mathsf{Z}$ such that, for any $z \in \mathsf{Z}$ and any neighborhood $\mathcal{N}$ of z_∞, $R^m(z, \mathcal{N}) > 0$ for at least one positive integer m.

(E-4) There exists a Markov kernel $\bar{R}$ on $(\mathsf{Z}^2 \times \{0,1\}, \mathcal{Z}^{\otimes 2} \otimes \mathcal{P}(\{0,1\}))$, a Markov kernel $\hat{R}$ on $(\mathsf{Z}^2, \mathcal{Z}^{\otimes 2})$, measurable functions $\bar{\alpha} : \mathsf{Z}^2 \to [0,1]$ and $W : \mathsf{Z}^2 \to [1, \infty)$ symmetric, and real numbers $(D, \zeta_1, \zeta_2, \rho) \in (\mathbb{R}_+)^3 \times (0,1)$ such that

for all $v = (z, z', u) \in \mathsf{X}^2 \times \{0,1\}$ and $n \geq 1$,

$$1 - \bar{\alpha} \leq \boldsymbol{\delta}_{\mathsf{Z}} \times W \quad \text{on } \mathsf{Z}^2 \,, \tag{4.9}$$

$$\forall z \in \mathsf{Z},\ \exists \gamma > 0,\ \sup\{W(z,z') : \boldsymbol{\delta}_{\mathsf{Z}}(z,z') < \gamma\} < \infty \,, \tag{4.10}$$

$$\begin{cases} \bar{R}(v; \cdot \times \mathsf{Z} \times \{0,1\}) = R(z, \cdot) \quad \text{and} \\ \bar{R}(v; \mathsf{Z} \times \cdot \times \{0,1\}) = R(z', \cdot) \,, \end{cases} \tag{4.11}$$

$$\bar{R}(v; \cdot \times \{1\}) = \bar{\alpha}(z,z')\, \hat{R}((z,z'), \cdot) \,, \tag{4.12}$$

$$\hat{R}^n((z,z'); \boldsymbol{\delta}_{\mathsf{Z}}) \leq D\rho^n \boldsymbol{\delta}_{\mathsf{Z}}(z,z') \,, \tag{4.13}$$

$$\hat{R}^n((z,z'); \boldsymbol{\delta}_{\mathsf{Z}} \times W) \leq D\rho^n \boldsymbol{\delta}_{\mathsf{Z}}^{\zeta_1}(z,z')\ W^{\zeta_2}(z,z') \,. \tag{4.14}$$

Based on these conditions, we can rely on the two following results. The existence of an invariant probability measure for R is given by the following result.

Lemma 9. *Under (E-1) and (E-2), R admits an invariant distribution π; moreover, $\pi V < \infty$.*

Proof. By [42, Theorem 2], Assumption (E-2) implies that the transition kernel R^m admits an invariant probability distribution denoted hereafter by π_m. Let $\tilde{\pi}$ be defined by, for all $A \in \mathcal{Z}$,

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

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

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

Letting $n \to \infty$, we then obtain $\tilde{\pi}(V \wedge M) \leq \frac{\beta}{1-\lambda} \wedge M$. Finally, by the monotone convergence theorem, letting $M \to \infty$, we get $\tilde{\pi} V < \infty$. □

Proposition 10. *Assume (E-1) (E-3) and (E-4). Then the Markov kernel R admits at most one unique invariant probability measure.*

Proof. This is extracted from the uniqueness part of the proof of [12, Theorem 6], see their Section 3. Note that our Condition (E-3) corresponds to their condition (A2) and our Condition (E-4) to their condition (A3) (their α, Q, $\bar{Q}$ and $Q^\sharp$ being our $\bar{\alpha}$, R, $\bar{R}$ and $\hat{R}$). □

Hence it is now clear that the conclusion of Theorem 6 holds if we can apply both Lemma 9 and Proposition 10. This is done according to the following successive steps.

Step 1 Under (A-3), the metric (3.5) makes $(\mathsf{Z}, \boldsymbol{\delta}_{\mathsf{Z}})$ locally compact, complete and separable, hence (E-1) holds true.

Step 2 Prove (E-2): this is done in Lemma 11 using (A-3), (A-5), (A-6) and the fact that, for all $y \in \mathsf{Y}$, $\psi\langle y \rangle$ is continuous on Z, as a consequence of $\mathrm{Lip}_1 < \infty$ in (A-4).

Step 3 Prove (E-3): this is done in Lemma 12, using (A-3), (A-4) and (A-7).

Step 4 Define $\bar{\alpha}$ and prove (4.9) and (4.10) in (E-4) with W as in (3.12): this directly follows from Conditions (A-8)(ii) and(A-8)(iii).

Step 5 Provide an explicit construction for $\bar{R}$ and $\hat{R}$ satisfying (4.11) and (4.12) in (E-4): this is done in Lemma 13 using (A-8)(i);

Step 6 Finally, we need to establish the additional properties of this $\hat{R}$ required in (E-4), namely, (4.13) and (4.14). This will be done in the final part of this section using the additional Lemma 14.

Let us start with **Step 2**.

Lemma 11. *If for all $y \in \mathsf{Y}$, $\psi\langle y\rangle$ is continuous on Z, then (A-5) and (A-6) imply (E-2).*

Proof. We first show that R is weak Feller, hence R^m is too, for any $m \geq 1$. Let $f : \mathsf{Z} \to \mathbb{R}$ be continuous and bounded. For all $z = (x_{(-p+1):0}, u_{(-q+1):(-1)}) \in \mathsf{Z}$, $Rf(z)$ is given by

$$\begin{aligned}\mathbb{E}_z\left[f(Z_1)\right] &= \mathbb{E}_z\left[f(x_{(-p+2):0}, X_1, u_{(-q+2):(-1)}, \Upsilon(Y_0))\right] \\ &= \int f(x_{(-p+2):0}, \psi\langle y\rangle(z), u_{(-q+2):(-1)}, \Upsilon(y))\, G(x_0; \mathrm{d}y)\ .\end{aligned}$$

Let us define $\tilde{f} : \mathsf{Z} \times \mathsf{Y} \to \mathbb{R}$ by setting, for $y \in \mathsf{Y}$ and z as above,

$$\tilde{f}(z, y) = f(x_{(-p+2):0}, \psi\langle y\rangle(z), u_{(-q+2):(-1)}, \Upsilon(y))\ .$$

Further define $\tilde{F} : \mathsf{Z} \times \mathsf{X} \to \mathbb{R}$ by setting, for all $z \in \mathsf{Z}$ and $x \in \mathsf{X}$,

$$\tilde{F}(z, x) = \int \tilde{f}(z, y)\, G(x; \mathrm{d}y)\ .$$

Hence, with these definitions, we have, for all $z \in \mathsf{Z}$, $Rf(z) = \tilde{F}(z, \Pi_p(z))$, and it is now sufficient to show that $\tilde{F}$ is continuous. We write, for all $z, z' \in \mathsf{Z}$ and $x, x' \in \mathsf{X}$, $\tilde{F}(z', x') - \tilde{F}(z, x) = A(z, z', x') + B(z, x, x')$ with

$$\begin{aligned}A(z, z', x') &= \int \left(\tilde{f}(z', y) - \tilde{f}(z, y)\right)\, G(x'; \mathrm{d}y) \\ B(z, x, x') &= \int \tilde{f}(z, y)\, \left(G(x'; \mathrm{d}y) - G(x; \mathrm{d}y)\right)\ .\end{aligned}$$

Since $z \mapsto \tilde{f}(z, y)$ is continuous for all y, we have $A(z, z', x') \to 0$ as $(z', x') \to (z, x)$ by (3.7) and dominated convergence. We have $B(z, x, x') \to 0$ as $x' \to x$, as a consequence of (A-5)(a) or of (A-5)(b). Hence $\tilde{F}$ is continuous and we have proved that R^m is weak Feller for all $m \in \mathbb{Z}_+^*$.

We now show that we can find $m \in \mathbb{Z}_+^*$, $\lambda \in (0, 1)$ and $\beta > 0$ such that $R^m V \leq \lambda V + \beta$ with $V : \mathsf{Z} \to \mathbb{R}_+$ defined by (3.9). We have, for all $n \geq q$,

$$\begin{aligned}R^n V(z) &= \mathbb{E}_z\left[\max_{0 \leq k < p} V_{\mathsf{X}}(X_{n-k}) \bigvee \max_{1 \leq k < q} \frac{V_{\mathrm{U}}(U_{n-k})}{|GV_{\mathsf{Y}}|_{V_{\mathsf{X}}}}\right] \\ &\leq \sum_{0 \leq k < p} \mathbb{E}_z\left[V_{\mathsf{X}}(X_{n-k})\right] + |GV_{\mathsf{Y}}|_{V_{\mathsf{X}}}^{-1} \sum_{1 \leq k < q} \mathbb{E}_z\left[V_{\mathsf{Y}}(Y_{n-k})\right] \\ &\leq 2 \sum_{0 \leq k < p \vee q} \mathbb{E}_z\left[V_{\mathsf{X}}(X_{n-k})\right]\ ,\end{aligned}$$

where we used that $V_{\mathrm{U}}(U_{n-k}) = V_{\mathrm{U}}(\Upsilon(Y_{n-k})) = V_{\mathsf{Y}}(Y_{n-k})$ and $\mathbb{E}_z\left[V_{\mathsf{Y}}(Y_{n-k})\right] = \mathbb{E}_z\left[GV_{\mathsf{Y}}(X_{n-k})\right] \leq |GV_{\mathsf{Y}}|_{V_{\mathsf{X}}} \mathbb{E}_z\left[V_{\mathsf{X}}(X_{n-k})\right]$, which is valid for $n - k \geq 0$. Now,

by (3.8), for any $\lambda \in (0,1)$, we can find $m \in \mathbb{Z}_+$ and $M > 0$ such that $\mathbb{E}_z\left[V_{\mathsf{X}}(X_{m-k})\right] \leq \lambda(V(z)+M)/(2(p\vee q))$ for $0 \leq k < p \vee q$. Hence $R^m V \leq \lambda V + \beta$ for $\beta = M\lambda$. □

We now proceed with **Step 3**.

Lemma 12. *(A-3), (A-4) and (A-7) imply (E-3).*

Proof. For any $y \in \mathsf{Y}$ and $n \in \mathbb{Z}_+ \cup \{\infty\}$, we denote by $c_n(y)$ the constant sequence with length n whose all entries are equal to y.

Let z, z' in Z. Denote, for all $n \in \mathbb{Z}_+^*$, $x^{(n)} = \psi^\theta\langle c_n(y_0)\rangle(z)$ and $\tilde{x}^{(n)} = \psi^\theta\langle c_n(y_0)\rangle(z')$. Then, for all $n \in \mathbb{Z}_+^*$,

$$\boldsymbol{\delta}_{\mathsf{X}}(x^{(n)}, \tilde{x}^{(n)}) \leq \mathrm{Lip}_n^\theta\, \boldsymbol{\delta}_{\mathsf{Z}}\left(z, z'\right) , \tag{4.15}$$

$$\boldsymbol{\delta}_{\mathsf{Z}}(x^{(n+1)}, x^{(n)}) \leq \mathrm{Lip}_n^\theta\, \boldsymbol{\delta}_{\mathsf{Z}}\left(z, \Psi_{y_0}^\theta(z)\right) . \tag{4.16}$$

By Lemma 15, the right-hand side in (4.16) is decreasing geometrically fast and $x^{(n)}$ converges to a point $\psi_\infty(y_0)$ which does not depend on z by (4.15). Set $z_\infty = (\psi_\infty(y_0), \ldots, \psi_\infty(y_0), \Upsilon(y_0), \ldots, \Upsilon(y_0)) \in \mathsf{Z}$. Then, by (4.4) and (3.5), for any $\delta > 0$, there exists $m \in \mathbb{Z}_+^*$ such that $\boldsymbol{\delta}_{\mathsf{Z}}(z_\infty, \Psi^\theta\langle c_m(y_0)\rangle(z)) < \delta$ and thus

$$\begin{aligned} R^m\left(z; \{z : \boldsymbol{\delta}_{\mathsf{Z}}(z_\infty, z) < \delta\}\right) &\geq \mathbb{P}_z(Z_m = \Psi^\theta\langle c_m(y_0)\rangle(z)) \\ &\geq \mathbb{P}_z(Y_k = y_0,\ \forall k \in \{1, \ldots, m\}) \\ &= \prod_{k=1}^m G^\theta\left(\psi^\theta\langle c_k(y_0)\rangle(z); \{y_0\}\right) . \end{aligned}$$

If (A-7) (a) holds, then (2.3) implies $G^\theta(x; \{y_0\}) > 0$ for all $x \in \mathsf{X}$, and we conclude that z_∞ is a reachable point.

If we now assume (A-7)(b), we then have that $y \mapsto \Psi^\theta\langle c_m(y)\rangle(z)$ is continuous on Y (by (4.1) and (4.2)) and thus there exists $\delta' > 0$ such that, for all y such that $\boldsymbol{\delta}_{\mathrm{U}}(\Upsilon(y), \Upsilon(y_0)) < \delta'$, $\boldsymbol{\delta}_{\mathsf{Z}}(z_\infty, \Psi^\theta\langle c_m(y)\rangle(z)) < 2\delta$, and thus

$$R^{m+1}\left(z; \{z : \boldsymbol{\delta}_{\mathsf{Z}}(z_\infty, z) < 2\delta\}\right) \geq \mathbb{P}_z(\boldsymbol{\delta}_{\mathrm{U}}(\Upsilon(Y_k), \Upsilon(y_0)) < \delta',\ 1 \leq k \leq m) .$$

Choose now y_0 such that $\nu(\{y : \boldsymbol{\delta}_{\mathrm{U}}(\Upsilon(y), \Upsilon(y_0)) < \delta'\}) > 0$ (there must be such a y_0 since U is separable by (A-3)) so that (2.3) implies $\mathbb{P}_z(\boldsymbol{\delta}_{\mathrm{U}}(\Upsilon(Y_\ell), \Upsilon(y_0)) < \delta' | \mathcal{F}_\ell) = G^\theta(X_\ell; \{y : \boldsymbol{\delta}_{\mathrm{U}}(\Upsilon(y), \Upsilon(y_0)) < \delta'\}) > 0$ $\mathbb{P}_z$-a.s. for all $k \geq 1$. It follows that for all $\ell = m, m-1, \ldots, 1$, conditioning on $\mathcal{F}_\ell$, $\mathbb{P}_z(\boldsymbol{\delta}_{\mathrm{U}}(\Upsilon(Y_k), \Upsilon(y_0)) < \delta',\ 1 \leq k \leq \ell) = 0$ implies $\mathbb{P}_z(\boldsymbol{\delta}_{\mathrm{U}}(\Upsilon(Y_k), \Upsilon(y_0)) < \delta',\ 1 \leq k \leq \ell - 1) = 0$. We conclude that $R^{m+1}\left(z; \{z : \boldsymbol{\delta}_{\mathsf{Z}}(z_\infty, z) < 2\delta\}\right) > 0$ and z_∞ is a reachable point. □

We now proceed with **Step 4**. Let us define $\bar{\alpha} = \alpha \circ \Pi_p^{\otimes 2}$. By (3.5) and (3.12), we have $W \geq W_{\mathsf{X}} \circ \Pi_p^{\otimes 2}$ and $\boldsymbol{\delta}_{\mathsf{Z}} \geq \boldsymbol{\delta}_{\mathsf{X}} \circ \Pi_p^{\otimes 2}$. Hence (4.9) follows from (A-8)(iii). Condition (4.10) directly follows from (A-8)(ii) and the definition of W in (3.12).

We now proceed with **Step 5**.

Lemma 13. *Let $\alpha : \mathsf{X}^2 \to [0,1]$ and $\phi : \mathsf{X}^2 \to \mathsf{X}$ be measurable functions satisfying (A-8)(i) and define the Markov kernel $\hat{R}$ on $(\mathsf{Z}^2, \mathcal{Z}^{\otimes 2})$ by*

$$\hat{R}f(v) = \int_{\mathsf{Y}} f \circ \Psi_y^{\otimes 2}(v)\, G\left(\phi \circ \Pi_p^{\otimes 2}(v); \mathrm{d}y\right) . \tag{4.17}$$

Then one can define a Markov kernel $\bar{R}$ on $(\mathsf{Z}^2 \times \{0,1\}, \mathcal{Z}^{\otimes 2} \otimes \mathcal{P}(\{0,1\}))$ which satisfies (4.11) and (4.12).

Proof. We first define a probability kernel $\bar{H}$ from Z^2 to $\mathcal{Y}^{\otimes 2} \otimes \mathcal{P}(\{0,1\})$ Let $(z, z') \in \mathsf{Z}^2$ and set $x = \Pi_p(z)$ and $x' = \Pi_p(z')$. We define $\bar{H}((z, z'); \cdot)$ as the distribution of (Y, Y', ϵ) drawn as follows. We first draw a random variable $\bar{Y}$ taking values in Y with distribution $G(\phi(x, x'); \cdot)$. Then we define (Y, Y', ϵ) by separating the two cases, $\alpha(x, x') = 1$ and $\alpha(x, x') < 1$.

- Suppose first that $\alpha(x, x') = 1$. Then by (A-8)(i), we have $G(x; \cdot) = G(x'; \cdot) = G(\phi(x, x'); \cdot)$. In this case, we set $(Y, Y', \epsilon) = (\bar{Y}, \bar{Y}, 1)$.
- Suppose now that $\alpha(z, z') < 1$. Then, using (3.11), the functions $(1 - \alpha(x, x'))^{-1} [g(x; \cdot) - \alpha(x, x') g(\phi(x, x'); \cdot)]$ and $(1 - \alpha(x, x'))^{-1} [g(x'; \cdot) - \alpha(x, x') g(\phi(x, x'); \cdot)]$ are probability density functions with respect to ν and we draw Λ and Λ' according to these two density functions, respectively. We then draw ϵ in $\{0, 1\}$ with mean $\alpha(x, x')$ and, assuming $\bar{Y}$, Λ, Λ' and ϵ to be independent, we set

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

One can easily check that the kernel $\bar{H}$ satisfies the following marginal conditions, for all $(z, z') \in \mathsf{Z}^2$ and $B \in \mathcal{Y}$,

$$\begin{cases} \bar{H}((z, z'); B \times \mathsf{Y} \times \{0, 1\}) = H(z; B) = G(\Pi_p(z); B), \\ \bar{H}((z, z'); \mathsf{Y} \times B \times \{0, 1\}) = H(z'; B) = G(\Pi_p(z'); B), \end{cases} \tag{4.18}$$

Define the Markov kernel $\bar{R}$ on $(\mathsf{Z}^2 \times \{0, 1\}, \mathcal{Z}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\}))$ by setting for all $(z, z', u) \in \mathsf{Z}^2 \times \{0, 1\}$ and $A \in \mathcal{Z}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\})$,

$$\bar{R}((z, z', u); A) = \int \mathbb{1}_A (\Psi_y(z), \Psi_{y'}(z'), u_1) \; \bar{H}((z, z'); \mathrm{d}y \; \mathrm{d}y' \; \mathrm{d}u_1).$$

Then (4.18) and (4.8) immediately gives (4.11). To conclude the proof we check (4.12). We have, for all $v = (z, z', u) \in \mathsf{Z}^2 \times \{0, 1\}$ and $A \in \mathcal{Z}^{\otimes 2}$,

$$\bar{R}(v; A \times \{1\}) = \mathbb{E} \left[\mathbb{1}_A \left(\Psi_y(\bar{Y}), \Psi_{y'}(\bar{Y}) \right) \mathbb{1}_{\{\epsilon = 1\}} \right] ,$$

where $\bar{Y}$ and ϵ are independent and distributed according to $G(\phi \circ \Pi_p^{\otimes 2}(z, z'); \cdot)$ and a Bernoulli distribution with mean $\bar{\alpha}(z, z')$. This, and the definition of $\hat{R}$ in (4.17) lead to (4.12). □

In order to achieve **Step 6**, we rely on the following result which is an adaption of [12, Lemma 9].

Lemma 14. *Assume that there exists* $(\varrho, D_1, \ell) \in (0, 1) \times \mathbb{R}_+ \times \mathbb{Z}_+^*$ *such that for all* $(z, z') \in \mathsf{Z}^2$,

$$\hat{R}\left((z, z'); \{\boldsymbol{\delta}_{\mathsf{Z}} \leq D_1 \boldsymbol{\delta}_{\mathsf{Z}}(z, z')\}\right) = 1, \tag{4.19}$$

$$\hat{R}^{\ell}\left((z, z'); \{\boldsymbol{\delta}_{\mathsf{Z}} \leq \varrho \boldsymbol{\delta}_{\mathsf{Z}}(z, z')\}\right) = 1, \tag{4.20}$$

and that $W : \mathsf{Z}^2 \to \mathbb{R}_+$ *satisfies*

$$\lim_{\zeta \to \infty} \limsup_{n \to \infty} \frac{1}{n} \; \ln \sup_{v \in \mathsf{Z}^2} \frac{\hat{R}^n W(v)}{W^{\zeta}(v)} \leq 0 \; , \tag{4.21}$$

Then, (4.13) *and* (4.14) *hold.*

Proof. Note that (4.19) implies, for any non-negative measurable function f on Z^2, $\hat{R}(\boldsymbol{\delta}_{\mathsf{Z}} \times f) \le D_1 \left(\boldsymbol{\delta}_{\mathsf{Z}} \times \hat{R} f\right)$ and (4.20) implies $\hat{R}^\ell(\boldsymbol{\delta}_{\mathsf{Z}} \times f) \le \varrho\,(\boldsymbol{\delta}_{\mathsf{Z}} \times \hat{R}^\ell f)$. Hence for all $n \ge 1$, writing $n = k\ell + r$, where $r \in \{0, \dots, \ell-1\}$ and $k \in \mathbb{Z}_+$, we get, setting $\rho' = \varrho^{1/\ell}$,

$$\hat{R}^n(\boldsymbol{\delta}_{\mathsf{Z}} \times f) \le D_1^r\, \varrho^k \left(\boldsymbol{\delta}_{\mathsf{Z}} \times \hat{R}^n f\right) \le \left(1 \vee D_1^{\ell-1}\right) \varrho^{-1}\, \rho'^{\,n} \left(\boldsymbol{\delta}_{\mathsf{Z}} \times \hat{R}^n f\right) .$$

Taking $f \equiv 1$, we get (4.13) for any $\rho \in [\rho', 1)$. To obtain (4.14), we take $f = W$ and observe that (4.21) implies that, for any $\delta > 1$, there exists $\zeta > 0$ such that $\sup(\hat{R}^n W / W^\zeta) = O(\delta^n)$. Choosing δ small enough to make $\rho'\delta < 1$, we get (4.14) with $\rho = \rho'\delta$, $\zeta_1 = 1$ and $\zeta_2 = \zeta$. □

We finally conclude **Step 6**. By Lemma 14, it remains to check Conditions (4.19) and (4.20), and (4.21). For all $\ell \ge 1$, we have $\hat{R}^\ell\left(v; \left\{\Psi\langle y_{1:\ell}\rangle^{\otimes 2}(v) : y_{1:\ell} \in \mathsf{Y}^\ell\right\}\right) = 1$. Using (5.1), we thus get

$$\hat{R}^\ell\left(v; \boldsymbol{\delta}_{\mathsf{Z}} \le \boldsymbol{\delta}_{\mathsf{Z}}(v)\left(\mathbb{1}_{\{\ell<q\}} \vee \max_{(\ell-p)_+<m\le\ell} \mathrm{Lip}_m\right)\right) = 1 ,$$

and (4.19) and (4.20) both follow from (A-4).

We now check (4.21). By (4.17) (3.10) and (3.12), for all $n \ge q$ and $v \in \mathsf{Z}^2$, $\hat{R}^n W(v)$ can be written as

$$\begin{aligned}
\hat{R}^n W(v) &= \hat{\mathbb{E}}_v\left[\max_{0\le k<p} W_{\mathsf{X}}(X_{n-k}, X'_{n-k}) \bigvee \max_{1\le k<q} \frac{W_{\mathsf{Y}}(Y_{n-k}) \vee W_{\mathsf{Y}}(Y'_{n-k})}{|GW_{\mathsf{Y}} \circ \phi|_{W_{\mathsf{X}}}}\right] \\
&\le \sum_{0\le k<p\vee q} \hat{\mathbb{E}}_v\left[W_{\mathsf{X}}(X_{n-k}, X'_{n-k}) + \frac{W_{\mathsf{Y}}(Y_{n-k}) + W_{\mathsf{Y}}(Y'_{n-k})}{|GW_{\mathsf{Y}} \circ \phi|_{W_{\mathsf{X}}}}\right] \\
&\le 3 \sum_{0\le k<p\vee q} \hat{\mathbb{E}}_v\left[W_{\mathsf{X}}(X_{n-k}, X'_{n-k})\right] ,
\end{aligned}$$

where we used, for $n - k \ge 0$, $\hat{\mathbb{E}}_v\left[W_{\mathsf{Y}}(Y_{n-k})\right] = \hat{\mathbb{E}}_v\left[W_{\mathsf{Y}}(Y'_{n-k})\right] = \hat{\mathbb{E}}_v\left[GW_{\mathsf{Y}} \circ \phi(X_{n-k}, X'_{n-k})\right]$. We directly get that (iv) implies (4.21). As for the two conditions in (v), the first one implies that, for any $\rho \in (0,1)$, there exists $m > 0$ and $\beta > 0$ such that $\hat{R}^m W \le \rho W + \beta$, and the second one that there exists $\tau \ge 1$ and $C > 0$ such that $\hat{R}^r W \le C\, W^\tau$ for all $r = 0, \dots, m-1$. Combining the two previous bounds, we obtain, for all $n = mk + r$ with $k \in \mathbb{Z}_+$ and $0 \le r < m$,

$$\hat{R}^n W \le \rho^k \hat{R}^r W + \beta/(1-\rho) \le \rho^k\, C\, W^\tau + \beta/(1-\rho) \le C_m\, W^\tau ,$$

where C_m is a positive constant, not depending on n. Hence we obtain (4.21) and the proof is concluded.

4.4. **Proof of Theorem 8.** We apply [14, Theorem 1] to the embedded ODM(1,1) with hidden variable space Z derived in Section 4.1. Note that our conditions (A-1) and (A-2) yield their condition (A-1) and (A-2) on the embedded model with

$$\bar{V}(x) = \max\{V_{\mathsf{X}}(x_k) : 1 \le k \le p\} , \quad x = x_{1:p} \in \mathsf{X}^p .$$

Let us briefly check that (B-2), (B-3) and (B-4) in [14] hold. Conditions (B-2) and (B-3) correspond to our (B-1) and (B-2), noting, for the latter one, that ψ_y^θ here corresponds to the Ψ_y^θ defined in (4.1), inherited from the embedding. As for (B-4) in [14], we have that (B-4)(i) and (B-4)(ii) corresponds to our (B-3)(i)

and (B-3)(ii), but with X_1 in (B-4) replaced by $\mathsf{Z}_1 = \mathsf{X}^{p-1} \times \mathsf{X}_1 \times \mathsf{Y}^{q-1}$ with the latter X_1 as in (B-3). Also our condition (A-4), by Lemma 15 and (5.1) imply (B-4)(iii) for some $\varrho \in (0,1)$ by setting $\bar{\psi}(z) = C\,\boldsymbol{\delta}_{\mathsf{Z}}(z^{(\mathrm{i})}, z)$ for some $C > 0$. This $\bar{\psi}$ is locally bounded, hence (B-4)(iv) holds. Condition (B-4)(v) follows (up to a multiplicative constant) from (B-3)(iii) by observing that, $z^{(\mathrm{i})}$ has constant first p entries and constant $q-1$ last entries,

$$\bar{\psi}(\Psi_y^\theta(z^{(\mathrm{i})})) = C\,\left(\boldsymbol{\delta}_{\mathsf{X}}(x_1^{(\mathrm{i})}, \psi^\theta\langle y\rangle(z^{(\mathrm{i})})) \vee \boldsymbol{\delta}_{\mathrm{U}}(\Upsilon(y_1^{(\mathrm{i})}), \Upsilon(y))\right) .$$

The remaining conditions (vi), (vii) and (viii) of (B-2) follow directly from (3.14), (v) and (vi) in (B-3). All the conditions of [14, Theorem 1] are checked and this result gives that, for all $\theta, \theta_\star \in \Theta$, $\tilde{\mathbb{P}}^{\theta_\star}$-a.s., $p^\theta(y|Y_{-\infty:0})$ defined as in (3.1) is well defined for all $y \in \mathsf{Y}$, that, if $\theta = \theta_\star$, it is the density of Y_1 given $Y_{-\infty:0}$ with respect to ν, and also that the MLE $\hat{\theta}_{z^{(\mathrm{i})},n}$ satisfies (3.3) with $\Theta^\star$ defined by (3.2). Finally, by [13, Theorem 3], we also obtain that $\Theta^\star = [\theta_\star]$ and the proof is concluded.

4.5. **Proof of Theorem 4.** We prove (i), (ii) and (iii) successively.

Proof of (i). We apply Theorem 6 with $V_{\mathsf{X}}(x) = \mathrm{e}^{\tau\,|x|}$ for some arbitrary $\tau > 0$. Note that Remark 2(1) give (I-1), which, by Remark 5(1), give (A-4). Remark 5(4) gives (A-5), and (A-7)(a) is trivial in this example. Hence, it only remains to show that (A-6), and (A-8) hold.

We start with (A-6), with $V_{\mathsf{X}}(x) = \mathrm{e}^{\tau\,|x|}$. We can further set $V_{\mathrm{U}}(u) = \mathrm{e}^{\tau\,|u|}$, hence $V_{\mathsf{Y}}(y) = \mathrm{e}^{\tau|\ln(1+y)|}$ and, by Lemma 18, we then have $GV_{\mathsf{Y}}(x) \le 2\mathrm{e}^{(1+x_+)\tau}$ so that $|GV_{\mathsf{Y}}|_{V_{\mathsf{X}}} \le 2\mathrm{e}^{\tau}$. With these definitions, (3.9) leads to

$$V(z) \ge (2\mathrm{e}^{\tau})^{-1}\,\mathrm{e}^{\tau\,|z|_\infty}, \qquad z \in \mathsf{Z} . \tag{4.22}$$

Now, to bound $\mathbb{E}_z[V_{\mathsf{X}}(X_n)]$ as n grows, we see $V_{\mathsf{X}}(X_n)$ as $\mathrm{e}^{\tau|\lambda(Z_n)|}$ with the specific $\lambda = \Pi_p$ and, for any linear form λ on Z, we look for a recursion relation applying to

$$\mathbb{E}_z\left[\mathrm{e}^{\tau|\lambda(Z_1)|}\right] = \mathbb{E}\left[\mathrm{e}^{\tau\left|\tilde{\lambda}_z(\ln(1+V))\right|}\right] ,$$

where $V \sim \mathcal{P}(\mathrm{e}^{\Pi_p(z)})$ and, for all $z = (x_{(-p+1):0}, y_{(-q+1):(-1)})$, $\tilde{\lambda}_z : \mathbb{R} \to \mathbb{R}$ is defined by

$$\tilde{\lambda}_z(y_0) = \lambda(x_{(-p+2):0}, \psi\langle y_0\rangle(z), y_{(-q+1):(-1)}, y_0) . \tag{4.23}$$

Observing that $\tilde{\lambda}_z$ is an affine function, of the form $\tilde{\lambda}_z(y) = \vartheta_0 + \vartheta y$, we can apply Lemma 18 with $\zeta = x_0$ (and the trivial bound $|\vartheta_0| \vee |\vartheta_0 + \vartheta\zeta_+| \le |\vartheta_0| \vee |\vartheta_0 + \vartheta\zeta|$) and obtain that

$$\mathbb{E}_z\left[\mathrm{e}^{\tau|\lambda(Z_1)|}\right] \le c\,\mathrm{e}^{\tau\left(\left|\lambda\circ\widehat{\Psi}\langle 0\rangle(z)\right|\vee\left|\lambda\circ\widehat{\Psi}\langle 1\rangle(z)\right|\right)} \le c\,\left[\mathrm{e}^{\tau\left|\lambda\circ\widehat{\Psi}\langle 0\rangle(z)\right|} + \mathrm{e}^{\tau\left|\lambda\circ\widehat{\Psi}\langle 1\rangle(z)\right|}\right] ,$$

where we set $c = 2\mathrm{e}^{\tau\omega + b_1}$ and, for $w = 0, 1$ and all $z \in \mathsf{Z}$, $\widehat{\Psi}\langle w\rangle(z)$ is defined by $\widehat{\Psi}\langle w\rangle(z) = (x_{(-p+2):0}, x_1, y_{(-q+2):-1}, y_0)$ with

$$\begin{aligned}
&x_{-p+k} = \Pi_k\,(z) \quad \text{for} \quad 1 \le k \le p\ ,\\
&y_{-q+k} = \Pi_{p+k}\,(z) \quad \text{for} \quad 1 \le k < q\ ,\\
&u_0 = wx_0 \quad \text{and} \quad u_{1-k} = \Upsilon(y_{1-k}) \quad \text{for} \quad 1 < k < q\ ,\\
&x_1 = \sum_{k=1}^{p} a_k\,x_{1-k} + \sum_{k=1}^{q} b_k\,u_k\ .
\end{aligned}$$

Defining for all $w = w_{0:n-1} \in \{0,1\}^n$, $\widehat{\Psi}\langle w\rangle = \widehat{\Psi}\langle w_{n-1}\rangle \circ \ldots \widehat{\Psi}\langle w_0\rangle$, we get

$$\mathbb{E}_z[V_{\mathsf{X}}(X_n)] \le c^n \sum_{w\in\{0,1\}^n} \mathrm{e}^{\tau\left|\Pi_p\circ\widehat{\Psi}\langle w\rangle(z)\right|} \ . \tag{4.24}$$

Now observe that, for all $w = w_{0:n+q-1} \in \{0,1\}^{n+q}$, we can define $\Pi_p \circ \widehat{\Psi}\langle w_{0:(n+q-1)}\rangle(z)$ as x_{n+q} obtained by adding to the previous recursive equations, for $1 \le k < n+q$,

$$u_k = w_k x_k \quad \text{and} \quad x_{k+1} = \sum_{j=1}^{p} a_j\, x_{k+1-j} + \sum_{j=1}^{q} b_j\, u_{k+1-j} \ .$$

Note that, in this recursion, we can replace u_{k+1-j} by $w_{k+1-j}x_{k+1-j}$ for $k+1-j \ge 0$, hence $x_{q:(n+q)}$ satisfies the recursion (2.15) and it follows that $\Pi_p \circ \widehat{\Psi}\langle w_{0:(n+q-1)}\rangle(z)$ can be expressed as

$$\widehat{\psi}\langle w_{q:(n+q-1)}\rangle\left(\left(\Pi_{p-(\ell-q)_+}\left(\widehat{\Psi}\langle w_{0:(q-\ell)_+}\rangle(z)\right)\right)_{1\le\ell\le p\vee q}\right) \ .$$

Hence Condition (2.16) implies that, for all $z \in \mathsf{Z}$,

$$\lim_{n\to\infty} \sup\left\{\left|\Pi_p \circ \widehat{\Psi}\langle w\rangle(z)\right| : w \in \{0,1\}^n\right\} = 0 \ .$$

By linearity of $z \mapsto \widehat{\Psi}\langle w\rangle(z)$, it follows that

$$\lim_{n\to\infty} \sup\left\{|z|_\infty^{-1}\left|\Pi_p \circ \widehat{\Psi}\langle w\rangle(z)\right| : w \in \{0,1\}^n,\ z \in \mathsf{Z}\setminus\{0\}\right\} = 0 \ . \tag{4.25}$$

Hence using (4.24), we finally obtain for a positive sequence $\{\rho_n : n \in \mathbb{Z}_+\}$,

$$\mathbb{E}_z[V_{\mathsf{X}}(X_n)] \le (2c)^n\ \mathrm{e}^{\tau\,\rho_n\,|z|_\infty} \ , \quad \text{with} \quad \lim_{n\to\infty}\rho_n = 0 \ ,$$

which, with (4.22), leads to, for all $z \in \mathsf{Z}$ and $M > 0$,

$$\frac{\mathbb{E}_z\left[V_{\mathsf{X}}(X_n)\right]}{M+V(z)} \le (2c)^n\ \min\left(M^{-1}\,\mathrm{e}^{\tau\,\rho_n\,|z|_\infty}, (2\mathrm{e}^{\tau})^{-1}\ \mathrm{e}^{\tau\,(\rho_n-1)\,|z|_\infty}\right) \ .$$

Let $C > 0$ be arbitrarily chosen. Using the first term and the second term in this min for $|z|_\infty \le C$ and $|z|_\infty > C$ respectively, we get that, for any n such that $\rho_n < 1$,

$$\limsup_{M\to\infty}\sup_{z\in\mathsf{Z}} \frac{\mathbb{E}_z\left[V_{\mathsf{X}}(X_n)\right]}{M+V(z)} \le (2c)^n\ \ (2\mathrm{e}^{\tau})^{-1}\ \mathrm{e}^{\tau\,(\rho_n-1)\,C} \to 0 \text{ as } C\to\infty \ .$$

Using that $\{\rho_n : n \in \mathbb{Z}_+\}$ converges to 0, this holds for n large enough and we get (3.8), and (A-6) holds.

We now turn to the proof of (A-8). We can apply Lemma 7 with $\mathsf{C} = \mathbb{R} = \mathsf{X}$ and $\mathsf{S} = \{1\}$, μ being the Dirac mass at point 1. For all $(x,y) \in \mathsf{X}\times\mathsf{Y}$, let $j(x) = \mathrm{e}^{-\mathrm{e}^x}$ and $h(x;u) = \frac{\mathrm{e}^{x(\mathrm{e}^u-1)}}{(\mathrm{e}^u-1)!}$, which h satisfies (H'-1). Hence Lemma 7 gives that (A-8)(i) holds with

$$\alpha(x,x') = \frac{\mathrm{e}^{-\mathrm{e}^{x\vee x'}}}{\mathrm{e}^{-\mathrm{e}^{x\wedge x'}}} = \mathrm{e}^{-\left|\mathrm{e}^x-\mathrm{e}^{x'}\right|} \quad \text{and} \quad \phi(x,x') = x\wedge x' \ , \qquad x,x' \in \mathsf{X} \ .$$

Now for all $x,x' \in \mathsf{X}$, we have

$$1-\alpha(x,x') = 1-\mathrm{e}^{-\left|\mathrm{e}^x-\mathrm{e}^{x'}\right|} \le \left|\mathrm{e}^x-\mathrm{e}^{x'}\right| \le \mathrm{e}^{|x|\vee|x'|}\,|x-x'| \ .$$

We thus obtain (A-8)(iii) by setting $W_{\mathsf{X}}(x,x') = \mathrm{e}^{|x|\vee|x'|}$, and (A-8)(ii) also follows by setting $W_{\mathrm{U}}(y) = \mathrm{e}^{|y|}$. Since W_{U} is V_{U} with $\tau = 1$, we already saw that $GW_{\mathsf{Y}}(x) \leq 2\mathrm{e}^{1+x_+}$, hence $|W_{\mathsf{Y}} \circ \phi|_{W_{\mathsf{X}}} \leq 2\mathrm{e}$, and (3.12) leads to, for all z, z' in Z,

$$W(z,z') \geq (2\mathrm{e})^{-1}\, \mathrm{e}^{|z|_\infty \vee |z'|_\infty} \ . \tag{4.26}$$

It now remains to prove either (A-8)(iv) or (A-8)(v), which both involve $\hat{\mathbb{E}}_z\left[W_{\mathsf{X}}(X_n, X_n')\right]$. We proceed as previously when we bounded $\mathbb{E}_z\left[\mathrm{e}^{\tau|\lambda(Z_1)|}\right]$. Let $\tau \geq 1$. For any linear function $\lambda : \mathsf{Z}^2 \to \mathsf{Z}^2$, we have, for all $v = (z,z') \in \mathsf{Z}^2$,

$$\begin{aligned}\hat{\mathbb{E}}_v\left[\mathrm{e}^{\tau|\lambda(Z_1)|\vee|\lambda(Z_1')|}\right] &\leq \hat{\mathbb{E}}_v\left[\mathrm{e}^{\tau|\lambda(Z_1)|}\right] + \hat{\mathbb{E}}_z\left[\mathrm{e}^{\tau|\lambda(Z_1')|}\right] \\ &= \mathbb{E}\left[\mathrm{e}^{\tau\left|\tilde{\lambda}_z(\ln(1+V))\right|}\right] + \mathbb{E}\left[\mathrm{e}^{\tau\left|\tilde{\lambda}_{z'}(\ln(1+V))\right|}\right] \ ,\end{aligned}$$

where $V \sim \mathcal{P}(\mathrm{e}^{\phi\circ\Pi_p^{\otimes 2}(v)})$ and $\tilde{\lambda}_z$ is defined by (4.23). By definition of ϕ above, we have $\phi \circ \Pi_p^{\otimes 2}(v) \leq \Pi_p(z), \Pi_p(z')$. Hence Lemma 18 with $\zeta = \phi \circ \Pi_p^{\otimes 2}(v)$ and $\zeta' = \Pi_p(z)$ and $\Pi_p(z')$ successively, we obtain, similarly as before for bounding $\mathbb{E}_z\left[\mathrm{e}^{\tau|\lambda(Z_1)|}\right]$, that for all $v = (z,z') \in \mathsf{Z}^2$,

$$\begin{aligned}\hat{\mathbb{E}}_v\left[\mathrm{e}^{\tau(|\lambda(Z_1)|\vee|\lambda(Z_1')|)}\right] &\leq c \sum_{z''=z,z'} \sum_{w=0,1} \mathrm{e}^{\tau\left|\lambda\circ\widehat{\Psi}\langle w\rangle(z'')\right|} \\ &\leq 2c \sum_{w=0,1} \mathrm{e}^{\tau\left(\left|\lambda\circ\widehat{\Psi}\langle w\rangle(z)\right|\vee\left|\lambda\circ\widehat{\Psi}\langle w\rangle(z')\right|\right)} \ .\end{aligned} \tag{4.27}$$

Taking $\lambda = \Pi_p$, and observing that $\Pi_p \circ \widehat{\Psi}\langle w\rangle$ is a linear form for $w = 0,1$, we get, setting $c_0 = \max_{w=0,1}\sup_{|z|_\infty\leq 1}|\Pi_p \circ \widehat{\Psi}\langle w\rangle(z)|$, for all $v = (z,z') \in \mathsf{Z}^2$,

$$\hat{\mathbb{E}}_v\left[W_{\mathsf{X}}^{\tau}(X_1, X_1')\right] \leq 4c\,\mathrm{e}^{\tau\, c_0\, |v|_\infty} \text{ with } |v|_\infty := |z|_\infty \vee |z'|_\infty \ . \tag{4.28}$$

Thus, with (4.26), the second condition of (A-8)(v) holds with $\tau' = \tau(c_0 \vee 1)$. To conclude, it is now sufficient to show that the first condition of (A-8)(v) also holds. Iterating (4.27) and taking $\tau = 1$ and $\lambda = \Pi_p$, we thus get, for all $n \in \mathbb{Z}_+$ and $v = (z,z') \in \mathsf{Z}^2$,

$$\begin{aligned}\hat{\mathbb{E}}_v\left[W_{\mathsf{X}}(X_n, X_n')\right] &= \hat{\mathbb{E}}_v\left[\mathrm{e}^{|\lambda(Z_n)|\vee|\lambda(Z_n')|}\right] \\ &\leq (4c)^n \max\left\{\mathrm{e}^{\left|\Pi_p\circ\widehat{\Psi}\langle w\rangle(z)\right|\vee\left|\Pi_p\circ\widehat{\Psi}\langle w\rangle(z')\right|} : w \in \{0,1\}^n\right\} \ .\end{aligned}$$

Applying (4.25) and (4.26), we get that, for all $v \in \mathsf{Z}^2$,

$$\frac{\hat{\mathbb{E}}_v\left[W_{\mathsf{X}}(X_n, X_n')\right]}{M + W(v)} \leq (4c)^n \min\left\{\frac{\mathrm{e}^{\rho_n\,|v|_\infty}}{M}, (2\mathrm{e})\,\mathrm{e}^{(\rho_n - 1)\,|v|_\infty}\right\} \ ,$$

where $\{\rho_n : n \in \mathbb{Z}_+\}$ is a positive sequence converging to 0. We now proceed as for proving (A-6) previously: the first term in the min tends to 0 as $M \to \infty$ uniformly over $|v|_\infty \leq C$ for any $C > 0$, while, if $\rho_n < 1$, the second one tends to zero as $|v|_\infty \to \infty$. Hence, for n large enough, we have

$$\lim_{M\to\infty}\sup_{v\in\mathsf{Z}^2}\frac{\hat{\mathbb{E}}_v\left[W_{\mathsf{X}}(X_n, X_n')\right]}{M + W(v)} = 0 \ ,$$

and (A-8)(v) follows, which concludes the proof. □

Proof of (ii). We apply Theorem 8. We have already shown that (A-4), (A-1) and (A-2) hold in the proof of Assertion (i), with $V_{\mathsf{X}}(x) = \mathrm{e}^{\tau|x|}$ for any $\tau > 0$. Assumptions (B-1) and (B-2) obviously hold for the log-linear Poisson GARCH model (see Remark 6 for the second one). It now only remains to show that, using V_{X} as above, (B-3) is also satisfied. We set $\mathsf{X}_1 = \mathsf{X}$ which trivially satisfies (B-3)(i). Condition (B-3)(ii) is also immediate (see Remark 7). Next, we look for an adequate $\bar{\phi}$, H and $C \geq 0$ so that (iii) (iv) (v) and (vi) hold in (B-3). We have, for all $\theta \in \Theta$ and $(x, x', y) \in \mathbb{R}^2 \times \mathbb{Z}_+$,

$$\begin{aligned}\left|\ln g^{\theta}(x;y) - \ln g^{\theta}(x';y)\right| &\leq |x - x'|\, \mathrm{e}^{x \vee x'}\, y \\ &\leq |x - x'|\, \mathrm{e}^{\left|x - x_1^{(\mathrm{i})}\right| \vee \left|x' - x_1^{(\mathrm{i})}\right|}\, y\, \mathrm{e}^{x_1^{(\mathrm{i})}} .\end{aligned}$$

We thus set $H(u) = \mathrm{e}^{x_1^{(\mathrm{i})}}\, u$, $C = 1$ and $\bar{\phi}(y) = A + B\, y$ for some adequate non-negative A and B so that (iii) (using Remark 6 and $\Upsilon(y) = \ln(1+y) \leq y$), (iv) and (v) follow. Then we have $G^{\theta}\bar{\phi}(x) \leq A + B\, \mathrm{e}^{x} \lesssim V_{\mathsf{X}}$, provided that we chose $\tau \geq 1$. This gives (vi) and thus (B-3) holds true, which concludes the proof. □

Proof of (iii). We apply [15, Theorem 8] in which:

- Condition (A-1) corresponds to our assumption (A-1) shown in Point (i) above;
- Condition (A-2) is readily checked for the Log Poisson Garch model;
- Condition (L-1) is the same as our Condition (I-1), which were already proved for showing Point (i) above;
- Condition (L-2) is the same as our Condition (I-2) which holds by assumption in (iii);
- Condition (3.11) corresponds to checking that

$$\int \ln^{+}(\ln(1+y))\, \pi_{\mathsf{Y}}^{\theta}(\mathrm{d}y) < \infty . \tag{4.29}$$

  We already checked that $G^{\theta} V_{\mathsf{Y}} \lesssim V_{\mathsf{X}}$ with $V_{\mathsf{Y}}(y) = (1+y)^{\tau}$, implying $\int (1+y)^{\tau} \pi_{\mathsf{Y}}^{\theta}(\mathrm{d}y) < \infty$. Hence (4.29) holds.

Thus [15, Theorem 8] gives that $[\theta_\star]$ reduced to $\{\theta_\star\}$ and Assertion (iii) follows from (ii). □

4.6. **Proof of Theorem 5.** We prove (i), (ii) and (iii) successively.

Proof of (i). As for the Log-linear Poisson Garch(p,q), we apply Theorem 6 this time with $V_{\mathsf{X}}(x) = x$. Note that, since $a_k, b_k \geq 0$ for all k, Condition (2.20) implies $\sum_{k=1}^{p} a_k < 1$, which implies (I-1) and thus (A-4) by Remark 5(1). Also as for the log-linear Poisson Garch(p,q) case, Remark 5(4) gives (A-5), and (A-7)(a) is trivially satisfied. Hence, again, we only have to show that (A-6), and (A-8) hold. Here Υ is the identity mapping.

We start with (A-6), with $V_{\mathsf{X}}(x) = x$. We can further set $V_{\mathsf{Y}}(y) = y$ since, we then have $GV_{\mathsf{Y}}(x) = rV_{\mathsf{X}}(x)$ so that $|GV_{\mathsf{Y}}|_{V_{\mathsf{X}}} = r$. With these definitions, (3.9) leads to

$$V(z) \geq (1+r)^{-1}\, |z|_{\infty}, \qquad z \in \mathsf{Z} . \tag{4.30}$$

Note that $V_{\mathsf{X}}(X_1) = \Pi_p(Z_1)$, and for all linear function $\lambda : \mathsf{Z} \to \mathsf{Z}$, we have, for all $z \in \mathsf{Z}$,

$$\mathbb{E}_z[\lambda(Z_1)] = \lambda\left(\mathbb{E}_z[Z_1]\right) = \lambda \circ \hat{\Psi}(z) ,$$

where, $z = (x_{(-p+1):0}, y_{(-q+1):(-1)}) \in \mathsf{Z}$,

$$\hat{\Psi}(z) = (x_{(-p+2):0}, \omega + \sum_{k=1}^{p} a_k \; x_{1-k} + \sum_{k=2}^{q} b_k \, y_{1-k} + b_1 r x_0, y_{(-q+2):(-1)}, r x_0) \; .$$

Using the notation introduced in Section 4.2, this linear mapping exactly correspond to the $(p+q-1) \times (p+q-1)$ matrix $\mathbf{A} + r\mathbf{b}\boldsymbol{\varepsilon}_p^T$. From Lemma 17, Condition (2.20) is equivalent to

$$|\lambda|_{\max}(\mathbf{A} + r\mathbf{b}\boldsymbol{\varepsilon}_p^T) < 1, \tag{4.31}$$

where, for any square matrix $\mathbf{M}$, $|\lambda|_{\max}(\mathbf{M})$ denote the spectral radius of $\mathbf{M}$. Thus we get with (4.30), for all $n \in \mathbb{Z}_+$ and $z \in \mathsf{Z}$,

$$\mathbb{E}_z[X_{n+1}] = \Pi_p \left(\mathbb{E}_z[Z_{n+1}]\right) = \Pi_p \circ (\mathbf{A} + r\mathbf{b}\boldsymbol{\varepsilon}_p^T)^n z \leq C \, \rho^n \, V(z) \; ,$$

where $C > 0$ and $\rho < 1$, and (A-6) is proved.

We conclude with the proof of (A-8). Let us apply Lemma 7 with $\mathsf{C} = (0, \infty) = \mathsf{X}$ and $\mathsf{S} = \{1\}$, μ being the Dirac mass at point 1, $j(x) = (1+x)^{-r}$ and $h(x;y) = \frac{\Gamma(r+y)}{y!\,\Gamma(r)} \left(\frac{x}{1+x}\right)^y$, which satisfies (H'-1). This leads to α and ϕ satisfying (A-8)(i) by setting, for all $x, x' \in \mathbb{R}_+$

$$\alpha(x,x') = \left(\frac{1 + x \wedge x'}{1 + x \vee x'}\right)^r \quad \text{and} \quad \phi(x,x') = x \wedge x'.$$

Next, for all $(x,x') \in \mathsf{Z}^2$, we have

$$1 - \alpha(x,x') = 1 - \left(\frac{1 + x \wedge x'}{1 + x \vee x'}\right)^r \leq (1 \vee r) \, |x - x'|$$

We thus obtain (A-8)(iii) by setting $W_{\mathsf{X}}(x,x') = (1 \vee r)$. All the other conditions of (A-8) are trivially satisfied in this case, taking $W_{\mathrm{U}} \equiv 1$. □

Proof of (ii). We apply Theorem 8. We have already shown that (A-4), (A-1) and (A-2) hold in the proof of Assertion (i), with $V_{\mathsf{X}}(x) = x$. Assumption (B-1) follows from (2.14), and (B-2) from Remark 6. It now only remains to show that, using V_{X} as above, (B-3) is also satisfied.

Since Θ is compact, we can find $\underline{\omega}$ and $\overline{r}$ such that $\omega \geq \underline{\omega}$ and $r \leq \overline{r}$ for all $\theta = (\omega, a_{1:p}, b_{1:q}, r) \in \Theta$. We set $\mathsf{X}_1 = [\underline{\omega}, \infty)$ which then satisfies (B-3)(i). Condition (B-3)(ii) follows from Remark 7.

Next, we look for an adequate $\bar{\phi}$, H and $C \geq 0$ so that (iii) (iv) (v) and (vi) hold in (B-3). For all $\theta \in \Theta$, $(x,x') \in \mathsf{X}_1$ and $y \in \mathbb{Z}_+$, we have

$$\begin{aligned} \left|\ln g^\theta(x;y) - \ln g^\theta(x';y)\right| &= |(r+y)[\ln(1+x') - \ln(1+x)] + y[\ln x - \ln x']| \\ &\leq \left[(r+y)(1+\underline{\omega})^{-1} + y\,\underline{\omega}^{-1}\right] \; |x - x'| \\ &\leq \left[\overline{r} + 2y\,\underline{\omega}^{-1}\right] \; |x - x'| \; . \end{aligned}$$

We set $C = 0$, $H(s) = s$ and $\bar{\phi}(y) = A + B\,y$ for some adequate non-negative A and B so that (iii) (using Remark 6 and $\Upsilon(y) = y$), (iv) and (v) follow. Then we have $G^\theta \ln^+ \bar{\phi}(x) \leq A + B\,r\,x \lesssim V_{\mathsf{X}}$. This gives (vi) and thus (B-3) holds true, which concludes the proof. □

Proof of (iii). The proof of this point is similar to the proof of Theorem 4(iii) in Section 4.5, except that Condition (4.29) is replaced by

$$\int \ln^+(|y|)\ \pi^\theta_{\mathsf{Y}}(\mathrm{d}y) < \infty\ . \tag{4.32}$$

We already checked that $G^\theta V_{\mathsf{Y}} \lesssim V_{\mathsf{X}}$ with $V_{\mathsf{Y}}(y) = y$, implying $\int y\, \pi^\theta_{\mathsf{Y}}(\mathrm{d}y) < \infty$. Hence (4.32) holds and the proof is concluded. □

5. Useful lemmas

The following result is used in the proofs of Lemma 12 and Theorem 8.

Lemma 15. *(A-4) implies that for all $\theta \in \Theta$, there exists $C > 0$ and $\rho \in (0,1)$ such that $\mathrm{Lip}^\theta_n \le C\ \rho^n$ for all $n \in \mathbb{Z}^*_+$.*

Proof. By (3.4), (3.5) and (4.4), we have, for all $n \in \mathbb{Z}^*_+$, using the convention $\mathrm{Lip}^\theta_m = 1$ for $m \le 0$,

$$\sup_{y\in\mathsf{Y}^n, v\in\mathsf{Z}^2} \frac{\boldsymbol{\delta}_{\mathsf{Z}} \circ \Psi^\theta\langle y\rangle^{\otimes 2}(v)}{\boldsymbol{\delta}_{\mathsf{Z}}(v)} \le \mathbb{1}_{\{n<q\}} \vee \left(\max_{0\le j<p} \mathrm{Lip}^\theta_{n-j}\right)\ . \tag{5.1}$$

Hence (A-4) implies that there exists $m \ge 1$ and $L \in (0,1)$ such that, for all $y \in \mathsf{Y}^{m+1}$, $\Psi^\theta\langle y\rangle$ is L-Lipschitz. Now observe that, by (4.3), for all $n = km + r$ with $k \ge 0$ and $0 \le r < m$, for all $y = y_{-n:0} \in \mathsf{Y}^{n+1}$, we can write $\Psi^\theta\langle y\rangle$ as

$$\Psi^\theta\langle y_{1-m:0}\rangle \circ \Psi^\theta\langle y_{1-2m:(-m)}\rangle \circ \cdots \circ \Psi^\theta\langle y_{(1-km):-(k-1)m}\rangle \circ \Psi^\theta\langle y_{-n:(-km)}\rangle\ ,$$

and in this composition, the k first functions are L-Lipschitz and the last one is $L' = 1 \vee \max\left\{\mathrm{Lip}^\theta_j : 0 < j \le m\right\}$-Lipschitz. Hence, for all $z, z' \in \mathsf{Z}$,

$$\boldsymbol{\delta}_{\mathsf{X}}(\psi^\theta\langle y\rangle(z), \psi^\theta\langle y\rangle(z')) \le \boldsymbol{\delta}_{\mathsf{Z}}(\Psi^\theta\langle y\rangle(z), \Psi^\theta\langle y\rangle(z')) \le L'\ L^k\, \boldsymbol{\delta}_{\mathsf{Z}}(z, z')\ .$$

Hence the result by setting $\rho = L^{1/m} \in (0,1)$. □

The following lemma is straightforward using standard algebra and is used numerously in various particular cases. Its proof is omitted.

Lemma 16. *The characteristic polynomial P of $\mathbf{A} + r\mathbf{b}\boldsymbol{\varepsilon}_p^T$ is given by*

$$\det\left(\lambda\mathbf{I} - (\mathbf{A} + r\mathbf{b}\boldsymbol{\varepsilon}_p^T)\right) = \lambda^{q-1}\left(\lambda^p - \sum_{k=1}^p a_k\ \lambda^{p-k}\right) - r\lambda^{p-1}\sum_{k=1}^q\ b_k\lambda^{q-k}\ .$$

The following lemma is used in the proof of Theorem 5.

Lemma 17. *Let $p, q \ge 1$, $(r, a_{1:p}, b_{1:q}) \in \mathbb{R}^{1+p+q}_+$. Then Condition (2.20) is equivalent to (4.31) and implies*

$$|\lambda|_{\max}(\mathbf{A}) < 1\ . \tag{5.2}$$

Proof. From Lemma 16, $|\lambda|_{\max}(\mathbf{A} + r\mathbf{b}\boldsymbol{\varepsilon}_p^T) < 1$ is equivalent to have that for all $z \in \mathbb{C}$,

$$|z| \ge 1 \Longrightarrow z^{p+q-1} - \sum_{k=1}^p a_k\ z^{p+q-k-1} - r\ \sum_{k=1}^q b_k\ z^{p+q-k-1} \ne 0\ ,$$

which can be rewritten as

$$|z| \leq 1 \Longrightarrow 1 - \sum_{k=1}^{p} a_k \; z^k - r \; \sum_{k=1}^{q} b_k \; z^k \neq 0 \; . \tag{5.3}$$

Setting $c_k = a_k + r \; b_k \geq 0$ for $k = 1, \ldots, p \wedge q$, $c_k = a_k$ if $q < k \leq p$ and $c_k = r b_k$ if $p < k \leq q$, it only remains to show that this implication is equivalent to the condition $\sum_{i=1}^{p} c_i < 1$. If $\sum_{k=1}^{p} c_k < 1$, then

$$|z| \leq 1 \Longrightarrow \left| \sum_{k=1}^{p \vee q} c_k z^k \right| \leq \sum_{k=1}^{p \vee q} c_k < 1 \; .$$

and (5.3) follows.

Note that $z \mapsto 1 - \sum_{k=1}^{p \vee q} c_k z^k$ is mapping $[0,1]$ to $[1 - \sum_{k=1}^{p \vee q} c_k, 1]$. Hence, if $\sum_{k=1}^{p \vee q} c_k \geq 1$, then $[0,1] \subset [1 - \sum_{k=1}^{p \vee q} c_k, 1]$, and $z \mapsto 1 - \sum_{k=1}^{p \vee q} c_k z^k$ must have a zero in $[0,1]$, which contradicts the implication (5.3).

We thus have shown the equivalence between (2.20) and (4.31) in the case $r = 1$, from which the general case $r \geq 0$ immediately follows.

Now, since (2.20) obviously implies $\sum_{k=1}^{p} a_k < 1$, which in turn is equivalent to (5.2) (case $r = 0$), we also get the claimed implication. □

The following lemma is used in the proof of Theorem 4.

Lemma 18. *Let $\vartheta \in \mathbb{R}$. Then, for all $\vartheta_0 \in \mathbb{R}$ and $\zeta \in \mathbb{R}$, if $U \sim \mathcal{P}(\mathrm{e}^{\zeta})$, then*

$$\mathbb{E}[(1+U)^{\vartheta}] \leq \mathrm{e}^{(1+\zeta_+)\,\vartheta_+} \tag{5.4}$$

$$\mathbb{E}[\mathrm{e}^{|\vartheta_0 + \vartheta \ln(1+U)|}] \leq 2\mathrm{e}^{\vartheta_+} \, \mathrm{e}^{|\vartheta_0| \vee |\vartheta_0 + \vartheta \, \zeta_+|} \tag{5.5}$$

$$\leq 2\mathrm{e}^{\vartheta_+} \, \mathrm{e}^{|\vartheta_0| \vee |\vartheta_0 + \vartheta \, \zeta'_+|} \quad \textit{for all} \quad \zeta' \geq \zeta \; . \tag{5.6}$$

Proof. We separate the proof of (5.4) in three different cases by specifying the bound (5.4) in each case.

Case 1 For all $\vartheta \leq 0$, we have $\mathbb{E}[(1+U)^{\vartheta}] \leq 1$.
Case 2 For all $\vartheta > 0$ and $\zeta < 0$, we have $\mathbb{E}[(1+U)^{\vartheta}] \leq \mathrm{e}^{\vartheta}$.
Case 3 For all $\vartheta > 0$ and $\zeta \geq 0$, we have $\mathbb{E}[(1+U)^{\vartheta}] \leq \mathrm{e}^{\vartheta} \, \mathrm{e}^{\zeta \vartheta}$.

The bound in Case 1 is obvious. The bound in Case 2 follows from

$$\mathbb{E}[(1+U)^{\vartheta}] \leq \mathbb{E}\left[\mathrm{e}^{\vartheta U}\right] = \mathrm{e}^{\mathrm{e}^{\zeta}(\vartheta - 1)} \; .$$

Finally, the bound in Case 3 follows from the following inequalities, valid for all $\vartheta > 0$ and $\zeta \geq 0$,

$$\mathbb{E}\left[\mathrm{e}^{-\zeta\vartheta}(1+U)^{\vartheta}\right] \leq \mathbb{E}\left[\left(1 + \mathrm{e}^{-\zeta} U\right)^{\vartheta}\right] \leq \mathbb{E}\left[\mathrm{e}^{\vartheta \mathrm{e}^{-\zeta} U}\right] = \mathrm{e}^{\mathrm{e}^{\zeta}(\vartheta \mathrm{e}^{-\zeta} - 1)} \leq \mathrm{e}^{\vartheta} \; .$$

Hence we get (5.4).

Let us now prove (5.5). Observe that

$$\begin{aligned} \mathbb{E}[\mathrm{e}^{|\vartheta_0 + \vartheta \ln(1+U)|}] &\leq \mathbb{E}[\mathrm{e}^{\vartheta_0 + \vartheta \ln(1+U)}] + \mathbb{E}[\mathrm{e}^{-\vartheta_0 - \vartheta \ln(1+U)}] \\ &= \mathrm{e}^{\vartheta_0} \mathbb{E}[(1+U)^{\vartheta}] + \mathrm{e}^{-\vartheta_0} \mathbb{E}[(1+U)^{-\vartheta}] \; . \end{aligned}$$

Then using (5.4), we get

$$\mathbb{E}[\mathrm{e}^{|\vartheta_0 + \vartheta \ln(1+U)|}] \leq 2\mathrm{e}^{\vartheta_+} \; \exp\left[(\vartheta_0 + (\zeta_+ \vartheta)_+) \vee (-\vartheta_0 + (\zeta_+ \vartheta)_-)\right]$$

We conclude (5.5) and (5.6) by observing that, for all $a, b \in \mathbb{R}$, $(a + b_+) \vee (-a + b_-) = a \vee (a+b) \vee (-a) \vee (-a-b) = |a| \vee |a+b|$. □

References

[1] Krishnan Bhaskaran, Antonio Gasparrini, Shakoor Hajat, Liam Smeeth, and Ben Armstrong. Time series regression studies in environmental epidemiology. *International journal of epidemiology*, page dyt092, 2013.

[2] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31:307–327, 1986.

[3] Tim Bollerslev. Glossary to arch (garch). Technical report, CREATES Research Paper, September 2008.

[4] P. Bougerol and N. Picard. Stationarity of garch processes and of some nonnegative time series. *J. Econometrics*, 52(1992):115 – 127, 1992.

[5] Vasiliki Christou and Konstantinos Fokianos. Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*, 35(1):55–78, 2014.

[6] Vasiliki Christou and Konstantinos Fokianos. On count time series prediction. *Journal of Statistical Computation and Simulation*, 85(2):357–373, 2015.

[7] DR Cox. Statistical analysis of time-series: some recent developments. *Scand. J. Statist.*, 8(2):93–115, 1981.

[8] Yan Cui and Fukang Zhu. A new bivariate integer-valued garch model allowing for negative cross-correlation. *TEST*, 27(2):428–452, Jun 2018.

[9] RA Davis, WTM Dunsmuir, and SB Streett. Observation-driven models for Poisson counts. *Biometrika*, 90(4):777–790, DEC 2003.

[10] R.A. Davis and H. Liu. Theory and inference for a class of observation-driven models with application to time series of counts. *Preprint, arXiv:1204.3915*, 2012.

[11] Richard A Davis, WTM Dunsmuir, and Y Wang. Modeling time series of count data. *Statistics textbooks and monographs*, 158:63–114, 1999.

[12] R. Douc, P. Doukhan, and E. Moulines. Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stoch. Proc. App.*, 123(7):2620–2647, 2013.

[13] R. Douc, F. Roueff, and T. Sim. The maximizing set of the asymptotic normalized log-likelihood for partially observed markov chains. *Annals of Applied Probability*, 26(4):2357–2383, 2016.

[14] Randal Douc, François Roueff, and Tepmony Sim. Handy sufficient conditions for the convergence of the maximum likelihood estimator in observation-driven models. *Lithuanian Mathematical Journal*, 55(3):367–392, 2015.

[15] Randal Douc, François Roueff, and Tepmony Sim. Necessary and sufficient conditions for the identifiability of observation-driven models. preprint, April 2019.

[16] Paul Doukhan, Konstantinos Fokianos, and Dag Tjøstheim. On weak dependence conditions for Poisson autoregressions. *Statist. Probab. Lett.*, 82(5):942–948, 2012.

[17] René Ferland, Alain Latour, and Driss Oraichi. Integer-valued GARCH process. *J. Time Ser. Anal.*, 27(6):923–942, 2006.

[18] K. Fokianos and D. Tjøstheim. Log-linear poisson autoregression. *J. of Multivariate Analysis*, 102(3):563–578, 2011.

[19] Konstantinos Fokianos, Michael H Neumann, et al. A goodness-of-fit test for poisson count processes. *Electronic Journal of Statistics*, 7:793–819, 2013.

[20] Konstantinos Fokianos, Anders Rahbek, and Dag Tjøstheim. Poisson autoregression. *J. Am. Statist. Assoc.*, 104(488):1430–1439, 2009. With electronic supplementary materials available online.

[21] Konstantinos Fokianos and Dag Tjøstheim. Nonlinear poisson autoregression. *Annals of the Institute of Statistical Mathematics*, 64(6):1205–1225, 2012.

[22] Christian Francq and Jean-Michel Zakoian. Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4):605–637, 2004.

[23] Christian Francq and Jean-Michel Zakoïan. A tour in the asymptotic theory of garch estimation. *Handbook of Financial Time Series*, pages 85–111, 2009.

[24] Christian Francq and Jean-Michel Zakoian. *GARCH models: structure, statistical inference and financial applications.* John Wiley & Sons, 2011.

[25] Jürgen Franke. *Weak dependence of functional INGARCH processes.* Technische Universität Kaiserslautern, Fachbereich Mathematik, 2010.

[26] Andréas Heinen. Modelling time series count data: an autoregressive conditional poisson model. *Available at SSRN 1117187*, 2003.

[27] S. G. Henderson, D.S. Matteson, and D.B. Woodard. Stationarity of generalized autoregressive moving average models. *Electronic Journal of Statistics*, 5:800–828, 2011.

[28] AR Ives, B Dennis, KL Cottingham, and SR Carpenter. Estimating community stability and ecological interactions from time-series data. *Ecological monographs*, 73(2):301–330, 2003.

[29] B. G. Leroux. Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.*, 40:127–143, 1992.

[30] Roman Liesenfeld and Jean-Francois Richard. Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of empirical finance*, 10(4):505–531, 2003.

[31] Alexander M Lindner. Stationarity, mixing, distributional properties and moments of garch (p, q)–processes. In *Handbook of financial time series*, pages 43–69. Springer, 2009.

[32] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Cambridge University Press, London, 2009.

[33] Michael H. Neumann. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli*, 17(4):1268–1284, NOV 2011.

[34] J. Pfanzagl. On the measurability and consistency of minimum contrast estimates. *Metrica*, 14:249–272, 1969.

[35] Robert S Pindyck and Daniel L Rubinfeld. *Econometric models and economic forecasts*, volume 4. Irwin/McGraw-Hill Boston, 1998.

[36] Tina Hviid Rydberg and Neil Shephard. Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics*, 1(1):2–25, 2003.

[37] Tepmony Sim. *Maximum likelihood estimation in partially observed Markov models with applications to time series of counts.* Theses, Télécom ParisTech, March 2016.

[38] D. Straumann and T. Mikosch. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *ArXiv Mathematics e-prints*, February 2007.

[39] S. Streett. *Some observation driven models for time series of counts.* PhD thesis, Colorado State University, Department of Statistics, 2000.

[40] U.S. Geology Survey. Earthquake hazards program.

[41] Dag Tjøstheim. *Count Time Series with Observation-Driven Autoregressive Parameter Dynamics*. Handbook of Discrete-Valued Time Series. London: Chapman & Hall, 2015.

[42] R. L. Tweedie. Invariant measures for Markov chains with no irreducibility assumptions. *Journal of Applied Probability*, 25:275–285, 1988.

[43] Chao Wang, Heng Liu, Jian-Feng Yao, Richard A Davis, and Wai Keung Li. Self-excited threshold poisson autoregression. *Journal of the American Statistical Association*, 109(506):777–787, 2014.

[44] Scott L Zeger. A regression model for time series of counts. *Biometrika*, 75(4):621–629, 1988.

[45] Fukang Zhu. A negative binomial integer-valued GARCH model. *J. Time Series Anal.*, 32(1):54–67, 2011.

Department of Foundation Year, Institute of Technology of Cambodia, 12156 Phnom Penh, Cambodia
E-mail address: tepmony.sim@itc.edu.kh

Département CITI, CNRS UMR 5157, Télécom SudParis, 91000 Évry, France
E-mail address: randal.douc@telecom-sudparis.eu

LTCI, Télécom ParisTech, Université Paris-Saclay, 75013 Paris, France
E-mail address: roueff@telecom-paristech.fr