

# General-order observation-driven models: Ergodicity and consistency of the maximum likelihood estimator

Tepmony Sim<sup>1</sup>, Randal Douc<sup>2</sup> and François Roueff<sup>3</sup>

<sup>1</sup>MIT Research Unit,  
Institute of Technology of Cambodia,  
120404 Phnom Penh, Cambodia  
e-mail: [tepmony.sim@itc.edu.kh](mailto:tepmony.sim@itc.edu.kh)

<sup>2</sup>Département CITI, CNRS UMR 5157,  
Télécom SudParis,  
Institut Polytechnique de Paris,  
91000 Évry, France  
e-mail: [randal.douc@telecom-sudparis.eu](mailto:randal.douc@telecom-sudparis.eu)

<sup>3</sup>LTCI, Télécom Paris,  
Institut Polytechnique de Paris,  
19 place Marguerite Perey,  
91120 Palaiseau, France  
e-mail: [roueff@telecom-paris.fr](mailto:roueff@telecom-paris.fr)

**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.

**MSC2020 subject classifications:** Primary 60J05, 62F12; secondary 62M05, 62M10. ■

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

Received December 2020.

## Contents

1	Introduction . . . . .	3350
2	Definitions and examples . . . . .	3352

2.1	Observation-driven model of order $(p, q)$ . . . . .	3352
2.2	Equivalence-class consistency and identifiability . . . . .	3356
2.3	Three examples . . . . .	3356
2.3.1	Log-linear Poisson GARCH model . . . . .	3356
2.3.2	NBIN-GARCH model . . . . .	3358
2.3.3	The PARX model . . . . .	3359
3	General results . . . . .	3364
3.1	Preliminaries . . . . .	3364
3.2	Ergodicity . . . . .	3365
3.3	Convergence of the MLE . . . . .	3369
4	Postponed proofs . . . . .	3370
4.1	Embedding into an observation-driven model of order $(1, 1)$ . . .	3370
4.2	Proof of <a href="#">Theorem 8</a> . . . . .	3372
4.3	Proof of <a href="#">Theorem 10</a> . . . . .	3378
4.4	Proof of <a href="#">Theorem 4</a> . . . . .	3378
4.5	Proof of <a href="#">Theorem 5</a> . . . . .	3382
4.6	Proof of <a href="#">Theorem 7</a> . . . . .	3384
5	Useful lemmas . . . . .	3389
	Acknowledgements . . . . .	3390
	References . . . . .	3391

## 1. Introduction

Since they were introduced in [8], observation-driven models have been receiving renewed interest in recent years. These models are widely applied in various fields ranging from economics (see [38]), environmental study (see [2]), epidemiology and public health study (see [47, 12, 20]), finance (see [33, 39, 23, 27]) and population dynamics (see [31]). The celebrated GARCH(1, 1) model introduced in [3] as well as most of the models derived from this one are typical examples of ODMs; see [4] for a list of some of them. A list of contributions on this class of models specifically dealing with discrete data includes [42, 10, 29, 20, 22, 28, 23, 30, 36, 11, 19, 13, 21, 6, 7, 9] and [15].

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 examples). The literature concerning the asymptotic theory of the MLE when the observed variable has Poisson distribution includes [22, 23, 24] and [46]. For a more general case where the model belongs to the class of one-parameter exponential ODMs, such as the Bernoulli, the exponential, the negative binomial (with known shape parameter) and the Poisson autoregressive models, the consistency and the asymptotic normality of the MLE have been

derived in [11]. 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 [13], where the observed process may admit various conditional distributions. This result has later been extended and refined in [15]. 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 [44].

Here, following others (e.g. [42, 29]), 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 and in particular we investigate the two problems listed here below.

- a) Provide a complete set of conditions for general order ODMs implying that the process is ergodic.
- b) Prove the consistency of the MLE. Under the assumption of well-specified models, this can be treated in two separate sub-problems.
  - 1) Prove that the MLE is equivalence-class consistent.
  - 2) Characterize the set of identifiable parameters.

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 [13, 15] to the embedded model. Yet the particular form of the embedded model does not fit the usual assumptions tailored for standard first-order ODMs. This is why, as pointed out in [44], solving Problem a) is not just a formal extension of available results. To obtain Theorem 8 where Problem a) is addressed, we derive conditions by 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). Once the ergodicity of the model is proved, we can solve Sub-problem b) 1). This is done in Theorem 10, which is obtained almost for free from the embedding in the ODM(1,1) case. Sub-problem b) 2) is much more involved and has been addressed in [17].

To demonstrate the generality, applicability and efficiency of our approach, we apply our results to three specific integer-valued ODMs, namely, the log-linear Poisson GARCH( $p, q$ ) model, the negative binomial integer-valued GARCH or the NBIN-GARCH( $p, q$ ) model and the Poisson AutoRegressive model with eXogenous variables, known as the PARX model. 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. For the PARX( $p, q$ ) model, which can be considered as a vector-valued ODM in our study, such results are available in [1] but our approach leads to significantly different assumptions, as will be shown in Section 2.3.3. Numerical experiments involving the log-linear Poisson

GARCH( $p, q$ ) and the NBIN-GARCH ( $p, q$ ) models can be found in [40, Section 5.5] for a set of earthquake data from Earthquake Hazards Program [43].

The paper is structured as follows. Definitions used throughout the paper are introduced in Section 2, where we also state our results on three 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  $\ell \leq m$ , 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 (V)LODM). Let  $(\mathsf{X}, \mathcal{X})$ ,  $(\mathsf{Y}, \mathcal{Y})$  and  $(\mathsf{U}, \mathcal{U})$  be measurable spaces, respectively called the *latent space*, the *observation space* and the *reduced observation space*. Let  $(\Theta, \Delta)$  be a compact metric space, called the *parameter space*. Let  $\Upsilon$  be a measurable function from  $(\mathsf{Y}, \mathcal{Y})$  to  $(\mathsf{U}, \mathcal{U})$ . Let  $\left\{ (x_{1:p}, u_{1:q}) \mapsto \tilde{\psi}_{u_{1:q}}^\theta(x_{1:p}) : \theta \in \Theta \right\}$  be a family of measurable functions from  $(\mathsf{X}^p \times \mathsf{U}^q, \mathcal{X}^{\otimes p} \otimes \mathcal{U}^{\otimes q})$  to  $(\mathsf{X}, \mathcal{X})$ , called the *reduced link functions* and let  $\{G^\theta : \theta \in \Theta\}$  be a family of probability kernels on  $\mathsf{X} \times \mathcal{Y}$ , called the *observation kernels*.

(i) A time series  $\{Y_k : k \geq -q + 1\}$  valued in  $\mathsf{Y}$  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  $\Upsilon$  and observation kernel  $G^\theta$  if there exists a process  $\{X_k : k \geq -p + 1\}$  on  $(\mathsf{X}, \mathcal{X})$  such that for all  $k \in \mathbb{Z}_{\geq 0}$ ,

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

where  $\mathcal{F}_k = \sigma(X_{(-p+1):k}, Y_{(-q+1):(k-1)})$  and  $U_j = \Upsilon(Y_j)$  for all  $j > -q$ .

(ii) We further say that this model is a *vector linearly observation-driven model of order ( $p, q, p', q'$ )* (shortened as VLDM( $p, q, p', q'$ )) if moreover for some  $p', q' \in \mathbb{Z}_{>0}$ ,  $\mathsf{X}$  is a closed subset of  $\mathbb{R}^{p'}$ ,  $\mathsf{U} \subset \mathbb{R}^{q'}$  and, for all  $x = x_{0:(p-1)} \in \mathsf{X}^p$ ,  $u = u_{0:(q-1)} \in \mathsf{U}^q$ , and  $\theta \in \Theta$ ,

$$\tilde{\psi}_u^\theta(x) = \omega(\theta) + \sum_{i=1}^p A_i(\theta) x_{p-i} + \sum_{i=1}^q B_i(\theta) u_{q-i}, \quad (2.2)$$

for some mappings  $\omega$ ,  $A_{1:p}$  and  $B_{1:q}$  defined on  $\Theta$  and valued in  $\mathbb{R}^{p'}$ ,  $(\mathbb{R}^{p' \times p'})^p$  and  $(\mathbb{R}^{p' \times q'})^q$ . In the case where  $p' = q' = 1$ , the VLDM( $p, q, p', q'$ ) is

simply called a *linearly observation-driven model of order*  $(p, q)$  (shortened as LODM $(p, q)$ ).

**Remark 1.** Let us comment briefly on this definition.

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

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

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

$$\psi_y^\theta(x) = \tilde{\psi}_{\Upsilon^{\otimes q}(y)}^\theta(x), \quad x \in \mathbf{X}^p, y \in \mathbf{Y}^q. \tag{2.3}$$

However by inserting the mapping  $\Upsilon$ , 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 or VLODMs use the form of reduced link function in (2.2) although they may use various mappings  $\Upsilon$ 's. This is the case for the GARCH, log-linear Poisson GARCH and NBIN-GARCH, see below, but also of the bivariate integer-valued GARCH model [9].

- (2) When  $p = q = 1$ , then the ODM $(p, q)$  defined by (2.1) reduces to the (first-order) ODM considered in [13] and [15]. 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) Note that in our definition,  $\Upsilon$  does not depend on  $\theta$ . In some cases, one can simplify the link function  $\tilde{\psi}^\theta$ , by allowing one to let  $\Upsilon$  depend on the unknown parameter. To derive identifiability conditions or to prove the consistency of the MLE, the dependence of the distribution with respect to the unknown parameter  $\theta$  is crucial. In contrast, proving that the process is ergodic is always done *for a given*  $\theta$  and it is thus possible to use our set of conditions to prove ergodicity with  $\Upsilon$  depending on  $\theta$ , hence reasoning for a given  $\theta$  and a given  $\Upsilon$ . Moreover our set of conditions (namely Conditions (A-3), (A-4), (A-5), (A-6), (A-7) and (A-8) in Theorem 8) depends on  $\Upsilon$  only through the image space  $U$ , which can usually be taken to be the same even in situations where  $\Upsilon$  may depend on  $\theta$ .

The GARCH model has been extensively studied, see, for example, [5, 25, 26, 34, 27] 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 [44]. Our results will apply to the GARCH case but we do not improve the existing results in this case. The real interest of our approach lies in being able to address general-order integer-valued ODMs for which existing results are scarce.

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

$$Z_k = (X_{(k-p+1):k}, U_{(k-q+1):(k-1)}) \in Z, \quad (2.4)$$

where we defined

$$Z = \mathcal{X}^p \times \mathcal{U}^{q-1} \quad \text{endowed with the } \sigma\text{-field } \mathcal{Z} = \mathcal{X}^{\otimes p} \otimes \mathcal{U}^{\otimes(q-1)}. \quad (2.5)$$

For all  $\theta \in \Theta$  and all probability distributions  $\xi$  on  $(Z, \mathcal{Z})$ , we denote by  $\mathbb{P}_\xi^\theta$  the distribution of  $\{X_k, Y_{k'} : k > -p, k' > -q\}$  satisfying (2.1) and  $Z_0 \sim \xi$ , with the usual notation  $\mathbb{P}_z^\theta$  in the case where  $\xi$  is a Dirac mass at  $z \in Z$ . We replace  $\mathbb{P}$  by  $\mathbb{E}$  to denote the corresponding expectations.

We always assume that the observation kernel is dominated by a  $\sigma$ -finite measure  $\nu$  on  $(Y, \mathcal{Y})$ , that is, for all  $\theta \in \Theta$ , there exists a measurable function  $g^\theta : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}$  written as  $(x, y) \mapsto g^\theta(x; y)$  such that for all  $x \in \mathcal{X}$ ,  $g^\theta(x; \cdot)$  is the density of  $G^\theta(x; \cdot)$  with respect to  $\nu$ . Then 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  $z \in Z$ , the corresponding conditional density function  $p^\theta(y_{0:n}|z)$  with respect to  $\nu^{\otimes(n+1)}$  under parameter  $\theta \in \Theta$  is given by

$$y_{0:n} \mapsto \prod_{k=0}^n g^\theta(x_k; y_k), \quad (2.6)$$

where, setting  $u_k = \Upsilon(y_k)$  for  $k = 0, \dots, n$ , the sequence  $x_{0:n}$  is defined through the initial conditions and recursion equations

$$\begin{cases} x_k = \Pi_{p+k}(z), & -p < k \leq 0, \\ u_k = \Pi_{p+q+k}(z), & -q < k \leq -1, \\ x_k = \tilde{\psi}_{u_{(k-q):(k-1)}}^\theta(x_{(k-p):(k-1)}), & 1 \leq k \leq n, \end{cases} \quad (2.7)$$

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

$$\psi^\theta(y_{0:(n-1)})(z) := x_n, \quad \text{with } x_n \text{ defined by (2.7)}. \quad (2.8)$$

Note that for  $n = 0$ ,  $y_{0:-1}$  is the empty sequence and (2.8) is replaced by  $\psi^\theta(\emptyset)(z) = x_0 = \Pi_p(z)$ . Given the initial condition  $Z_0 = z^{(i)}$ , the (conditional) maximum likelihood estimator  $\hat{\theta}_{z^{(i)}, n}$  of the parameter  $\theta$  is thus defined by

$$\hat{\theta}_{z^{(i)}, n} \in \operatorname{argmax}_{\theta \in \Theta} \mathbf{L}_{z^{(i)}, n}^\theta, \quad (2.9)$$

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

$$L_{z^{(i)},n}^\theta := n^{-1} \sum_{k=1}^n \ln g^\theta \left( \psi^\theta \langle Y_{0:(k-1)} \rangle (z^{(i)}); Y_k \right). \tag{2.10}$$

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  $z^{(i)}$ ). We also use arbitrary values for the last  $q - 1$  entries of  $z^{(i)}$ . Note that we index our set of observations as  $Y_{0:n}$  (hence assuming  $1 + n$  observations to compute  $L_{z^{(i)},n}^\theta$ ). The iterated function  $\psi^\theta \langle Y_{0:k} \rangle$  can be cumbersome but is very easy to compute in practice using the recursion (2.7). Moreover, the same kind of recursion holds for its derivatives with respect to  $\theta$ , allowing one to apply gradient steps to locally maximize the likelihood.

In this contribution, we investigate the convergence of  $\hat{\theta}_{z^{(i)},n}$  as  $n \rightarrow \infty$  for some (well-chosen) value of  $z^{(i)}$  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  $((X \times Y)^\mathbb{Z}, (\mathcal{X} \times \mathcal{Y})^{\otimes \mathbb{Z}})$  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  $Y^\mathbb{Z}$ .

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  $\pi_X^\theta$  and  $\pi_Y^\theta$  the marginal distributions of  $X_0$  and  $Y_0$  under  $\mathbb{P}^\theta$ , on  $(X, \mathcal{X})$  and  $(Y, \mathcal{Y})$ , respectively. As a byproduct of the proof of (A-1), one usually obtains a function  $V_X : X \rightarrow \mathbb{R}_{\geq 0}$  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  $\theta \in \Theta$ ,  $\pi_X^\theta(V_X) < \infty$ .

It is here stated as an assumption for convenience. Note also that, in the following, for  $V : X \rightarrow \mathbb{R}_{\geq 0}$  and  $f : X \rightarrow \mathbb{R}$ , we denote the  $V$ -norm of  $f$  by

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

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

## 2.2. Equivalence-class consistency and identifiability

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  $\Theta$  and the corresponding equivalence class of  $\theta$  is denoted by  $[\theta] := \{\theta' \in \Theta : \theta \sim \theta'\}$ .

The equivalence relationship  $\sim$  was introduced by [32] as an alternative to the classical identifiability condition. Namely, we say that  $\hat{\theta}_{z^{(i)},n}$  is *equivalence-class consistent* at the true parameter  $\theta_*$  if

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{z^{(i)},n}, [\theta_*]) = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.} \quad (2.11)$$

Recall that  $\Delta$  is the metric endowing the parameter space  $\Theta$ . Therefore, in (2.11),  $\Delta(\hat{\theta}_{z^{(i)},n}, [\theta_*])$  is the distance between the MLE and the set of parameters having the same stationary distribution as the true parameter, and the convergence is considered under this common stationary distribution.

Identifiability can then be treated as a separate problem which consists in determining all the parameters  $\theta_*$  for which  $[\theta_*]$  reduces to the singleton  $\{\theta_*\}$ , so that equivalence-class consistency becomes the usual consistency at and only at these parameters. The identifiability problem is treated in [17, Proposition 11 and Theorem 12] for general order ODMs, where many cases of interest are specifically detailed. We will use in particular [17, Theorem 17 and Section 5.5].

## 2.3. Three examples

To motivate our general results we first make explicit three models of interest, namely, the log-linear Poisson GARCH( $p, q$ ), the NBIN-GARCH( $p, q$ ) and the PARX( $p, q$ ) 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. Such results are available for PARX model in [1] but our approach leads to significantly different assumptions, as will be shown in Section 2.3.3.

Once the ergodicity of the model is established, we can investigate the consistency of the MLE. This is done by first investigating the equivalence-class consistency of the MLE and then we only need to add an identifiability condition to get the consistency of the MLE, see Theorems 4, 5 and 7 hereafter. Such results pave the way for investigating the asymptotic normality. This is done in [40, Proposition 5.4.7(iii) and Proposition 5.4.15] for the first two examples and in [1, Theorems 2 and 3] for the third one.

### 2.3.1. Log-linear Poisson GARCH model

The Log-linear Poisson GARCH model is defined as follows in our setting.



**Example 1.** The Log-linear Poisson GARCH( $p, q$ ) Model parameterized by  $\theta = (\omega, a_{1:p}, b_{1:q}) \in \Theta \subset \mathbb{R}^{1+p+q}$  is a LODM( $p, q$ ) with affine reduced link function of the form (2.2) with coefficients given by

$$\begin{aligned} \omega(\theta) &= \omega, & A_i(\theta) &= a_i \quad \text{for } 1 \leq i \leq p \\ \text{and } B_i(\theta) &= b_i \quad \text{for } 1 \leq i \leq q, \end{aligned} \tag{2.12}$$

with observations space  $\mathbb{Y} = \mathbb{Z}_{\geq 0}$ , hidden variables space  $\mathbb{X} = \mathbb{R}$ , admissible mapping  $\Upsilon(y) = \ln(1+y)$ , and observation kernel  $G^\theta(x; \cdot)$  defined as the Poisson distribution with mean  $e^x$ .

Our condition for showing the ergodicity of general order log-linear Poisson GARCH models requires the following definition. For all  $x = x_{(1-(p \vee q)):0} \in \mathbb{R}^{p \vee q}$ ,  $m \in \mathbb{Z}_{\geq 0}$ , and  $w = w_{(-q+1):m} \in \{0, 1\}^{q+m}$ , define  $\widehat{\psi}^\theta(w)(x)$  as  $x_{m+1}$  obtained by the recursion

$$x_k = \sum_{j=1}^p a_j x_{k-j} + \sum_{j=1}^q b_j w_{k-j} x_{k-j}, \quad 1 \leq k \leq m+1 \tag{2.13}$$

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

**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

$$\lim_{m \rightarrow \infty} \max \left\{ \left| \widehat{\psi}^\theta(w)(x) \right| : w \in \{0, 1\}^{q+m} \right\} = 0 \quad \text{for all } x \in \mathbb{R}^{p \vee q}, \tag{2.14}$$

and define

$$P_p(z; a_{1:p}) = z^p - \sum_{k=1}^p a_k z^{p-k} \quad \text{and} \quad Q_q(z; b_{1:q}) = \sum_{k=0}^{q-1} b_{k+1} z^{q-1-k}. \tag{2.15}$$

Then, the following assertions hold.

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

$$V_X(x) = e^{\tau|x|}, \quad x \in \mathbb{R}. \tag{2.16}$$

(ii) For any  $x_1^{(i)} \in \mathbb{R}$  and  $y_1^{(i)} \in \mathbb{Z}_{\geq 0}$ , setting  $z^{(i)} = (x_1^{(i)}, \dots, x_1^{(i)}, \Upsilon(y_1^{(i)}), \dots, \Upsilon(y_1^{(i)})) \in \mathbb{R}^p \times \mathbb{R}^{q-1}$ , the MLE  $\hat{\theta}_{z^{(i)}, n}$  as defined by (2.9) is equivalence-class consistent, that is, (2.11) holds for any  $\theta_* \in \Theta$ .

(iii) If the true parameter  $\theta_* = (\omega^*, a_{1:p}^*, b_{1:q}^*)$  moreover satisfies that the polynomials  $P_p(\cdot; a_{1:p}^*)$  and  $Q_q(\cdot; b_{1:q}^*)$  defined by (2.15) have no common complex roots, then the MLE  $\hat{\theta}_{z^{(i)}, n}$  is consistent.

The proof is postponed to [Section 4.4](#).

**Remark 2.** Let us provide some insights about Condition [\(2.14\)](#).

- (1) Using the two possible constant sequences  $w$ ,  $w_k = 0$  for all  $k$  or  $w_k = 1$  for all  $k$  in [\(2.13\)](#), we easily see that [\(2.14\)](#) 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  $p < k \leq q$  and  $b_k = 0$  for  $q < k \leq p$  and where

$$\mathcal{S}_p = \left\{ c_{1:p} \in \mathbb{R}^p : \forall z \in \mathbb{C}, |z| \leq 1 \text{ implies } 1 - \sum_{k=1}^p c_k z^k \neq 0 \right\}. \quad (2.17)$$

- (2) A sufficient condition to have [\(2.14\)](#) is

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

Indeed, defining  $\rho$  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  $x \in \mathbb{R}^{p \vee q}$ ,

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

- (3) The first iteration [\(2.13\)](#) implies, for all  $w \in \{0, 1\}^q$  and  $x \in [-1, 1]^{p \vee q}$ ,

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

Hence a sufficient condition to have [\(2.18\)](#) (and thus [\(2.14\)](#)) is

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

- (4) When  $p = q = 1$ , by Points [\(1\)](#) and [\(3\)](#) above, Condition [\(2.14\)](#) is equivalent to have  $|a_1| < 1$  and  $|a_1 + b_1| < 1$ . This condition is weaker than the one derived in [\[13\]](#) where  $|b_1| < 1$  is also imposed.

### 2.3.2. NBIN-GARCH model

Our next example is the NBIN-GARCH( $p, q$ ), which is defined as follows in our setting.

**Example 2.** The NBIN-GARCH( $p, q$ ) model is a LODM( $p, q$ ) parameterized by  $\theta = (\omega, a_{1:p}, b_{1:q}, r) \in \Theta \subset \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0}^{p+q} \times \mathbb{R}_{>0}$  with affine reduced link function of the form [\(2.2\)](#) with coefficients given by [\(2.12\)](#), observations space  $\mathbb{Y} = \mathbb{Z}_{\geq 0}$ , hidden variables space  $\mathbb{X} = \mathbb{R}_{\geq 0}$ , admissible mapping  $\Upsilon(y) = y$ , and

observation kernels  $G^\theta(x; \cdot)$  defined as the negative binomial distribution with shape parameter  $r > 0$  and mean  $rx$ , that is, for all  $y \in \mathbb{Z}_{\geq 0}$ ,

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

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  $V_X(x) = x$  for all  $x \in \mathbb{R}_+$ .
- (ii) For any  $x_1^{(i)} \in \mathbb{R}_{>0}$  and  $y_1^{(i)} \in \mathbb{Z}_{\geq 0}$ , setting  $z^{(i)} = (x_1^{(i)}, \dots, x_1^{(i)}, y_1^{(i)}, \dots, y_1^{(i)}) \in \mathbb{R}_{\geq 0}^p \times \mathbb{Z}_{\geq 0}^{q-1}$ , the MLE  $\hat{\theta}_{z^{(i)}, n}$  as defined by (2.9) is equivalence-class consistent, that is, (2.11) holds for any  $\theta_* \in \Theta$ .
- (iii) If the true parameter  $\theta_* = (\omega^*, a_{1:p}^*, b_{1:q}^*, r^*)$  moreover satisfies that the polynomials  $P_p(\cdot; a_{1:p}^*)$  and  $Q_q(\cdot; b_{1:q}^*)$  defined by (2.15) have no common complex roots, then the MLE  $\hat{\theta}_{z^{(i)}, n}$  is consistent.

The proof is postponed to Section 4.5.

**Remark 3.** Clearly, under the setting of Example 2, if Eq. (2.1) has a stationary solution such that  $\mu = \int x \pi_X(dx) < \infty$ , taking the expectation on both sides of the second equation in (2.1) and using that  $\int y \pi_Y(dy) = r \int x \pi_X(dx)$ , 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 [48, 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.

### 2.3.3. The PARX model

The PARX model is similar to the standard INGARCH( $p, q$ ) model but with additional exogenous variables in the linear link function for generating the hidden variables. Following [1], these exogenous variables are assumed to satisfy some Markov dynamic of order 1 independently of the observations and of the hidden variables (see their Assumption 1). This leads to the following definition.

**Definition 6** (PARX model). Let  $d, p, q$  and  $r$  be positive integers. For  $\gamma = \gamma_{1:d} \in \mathbb{R}_{\geq 0}^d$ , we set  $\mathbf{f}_\gamma : \xi = (\xi_1, \dots, \xi_d) \mapsto \sum_{j=1}^d \gamma_j f_j(\xi_j)$  where  $f_i : \mathbb{R}^r \rightarrow \mathbb{R}_{\geq 0}$  are given functions for all  $i \in \{1, \dots, d\}$ . Let  $\mathcal{P}(x)$  denote the Poisson distribution with parameter  $x$  and  $L$  be a given Markov kernel on  $\mathbb{R}^r \times \mathcal{B}(\mathbb{R}^r)$ . We define the Poisson AutoRegressive model with eXogenous variables, shortly written as PARX( $p, q$ ), as a class of processes parameterized by  $\theta = (\omega, a_{1:p}, b_{1:q}, \gamma_{1:d}) \in \Theta$ , where  $\Theta$  is a compact subset of  $\mathbb{R}_{>0} \times \mathbb{R}_{\geq 0}^{p+q+d}$ , and satisfying

$$\begin{aligned} X_t &= \omega + \sum_{i=1}^p a_i X_{t-i} + \sum_{i=1}^q b_i Y_{t-i} + \mathbf{f}_\gamma(\Xi_{t-1}) \\ (Y_t, \Xi_t) | \mathcal{F}_t &\sim \mathcal{P}(X_t) \otimes L(\Xi_{t-1}; \cdot), \end{aligned} \quad (2.21)$$

where  $\mathcal{F}_t = \sigma(X_{(-p+1):t}, Y_{(-q+1):(t-1)}, \Xi_{(-q+1):(t-1)})$ .

**Remark 4.** Note that our  $a_1, \dots, a_p$  and  $b_1, \dots, b_q$  correspond to  $\beta_1, \dots, \beta_q$  and  $\alpha_1, \dots, \alpha_p$  of [1]. Here we allowed  $r$  to be different from  $d$  whereas in [1] it is assumed that  $d = r$  and they are denoted by  $d_x$ .

Since the exogenous variables are observed we can recast the PARX model into a VLDM( $p, q$ ) by including the exogenous variables into the observations  $Y_t$  and the hidden variable  $X_t$ . Namely, we can formally see the above PARX model as a special case of VLDM as follows.

**Example 3** (PARX model as a VLDM). Consider a PARX model as in Definition 6. Set  $\bar{Y}_t = (Y_t, \Xi_t)$ ,  $\bar{X}_t = (X_t, \Xi_{t-1})$ , which are valued in  $\bar{Y} = \mathbb{Z}_{\geq 0} \times \mathbb{R}^r$  and  $\bar{X} = \mathbb{R}_{\geq 0} \times \mathbb{R}^r$ . Then  $\{\bar{Y}_k : k \geq -q + 1\}$  is a VLDM( $p, q, 1 + r, 1 + d + r$ ) by setting

- the admissible mapping as  $\Upsilon(\bar{y}) := (y, f_1(\xi), \dots, f_d(\xi), \xi) \in \bar{U} := \mathbb{Z}_{\geq 0} \times \mathbb{R}_{\geq 0}^d \times \mathbb{R}^r$  for all  $\bar{y} = (y, \xi) \in \mathbb{Z}_{\geq 0} \times \mathbb{R}^r$ ;
- for all  $\theta = (\omega, a_{1:p}, b_{1:q}, \gamma_{1:d})$ ,  $\omega(\theta) := \begin{bmatrix} \omega \\ 0_{r,1} \end{bmatrix}$ ,  $A_k(\theta) := \begin{bmatrix} a_k & 0_{1,r} \\ 0_{r,1} & 0_{r,r} \end{bmatrix}$   
for  $k = 1, \dots, p$ ,  $B_1(\theta) := \begin{bmatrix} b_1 & [\gamma_1 \ \cdots \ \gamma_d] & 0_{1,r} \\ 0_{r,1} & 0_{r,d} & \mathbf{I}_r \end{bmatrix}$  and  $B_k(\theta) := \begin{bmatrix} b_k & 0_{1,d+r} \\ 0_{r,1} & 0_{r,d+r} \end{bmatrix}$  for  $k = 1, \dots, q$ ;
- for all  $\bar{x} = (x, \xi) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^r$ ,  $G^\theta(\bar{x}; \cdot) := \mathcal{P}(x) \otimes L(\xi; \cdot)$ .

We end up this section by stating our result for general PARX model. For establishing the ergodicity of the PARX model, we need some assumptions on the dynamic of the exogeneous variables through the Markov kernel  $L$ . Namely we consider the following assumptions on the kernel  $L$ .

(L-1) The Markov kernel  $L$  admits a kernel density  $\ell$  with respect to some measure  $\nu_L$  on Borel sets of  $\mathbb{R}^r$  such that

- for all  $(\xi, \xi') \in \mathbb{R}^r \times \mathbb{R}^r$ ,  $\ell(\xi; \xi') > 0$ ;

(ii) for all  $\xi \in \mathbb{R}^r$ , there exists  $\delta > 0$  such that

$$\int_{\mathbb{R}^r} \sup_{\|\xi' - \xi\| \leq \delta} \ell(\xi'; \xi'') \nu_L(d\xi'') < \infty,$$

where  $\|\cdot\|$  denotes the Euclidean norm on  $\mathbb{R}^r$ .

(L-2) The Markov kernel  $L$  is weak-Feller ( $Lf$  is bounded and continuous for any  $f$  that is bounded and continuous).

(L-3) There exist a probability measure  $\pi_L$  on Borel sets of  $\mathbb{R}^r$ , a measurable function  $V_L : \mathbb{R}^r \rightarrow [1, \infty)$  and constants  $(\varrho, C) \in (0, 1) \times \mathbb{R}_{\geq 0}$  such that

- (i)  $L$  is  $\pi_L$ -invariant,
- (ii)  $\pi_L(V_L) < \infty$  and  $\{V_L \leq M\}$  is a compact set for any  $M > 0$ ,
- (iii) for all  $i \in \{1, \dots, d\}$ , we have  $f_i \lesssim V_L$ .
- (iv) for all Borel functions  $h : \mathbb{R}^r \rightarrow \mathbb{R}$  such that  $h \lesssim V_L$ , we have for all  $(\xi, n) \in \mathbb{Z}_{\geq 0}$ ,

$$|L^n h(\xi) - \pi_L(h)| \leq C \varrho^n |h|_{V_L} V_L(\xi).$$

(L-4) There exists a constant  $M \geq 1$  such that for all  $(\xi, \xi') \in \mathbb{R}^{2r}$ ,

$$d_{TV}(L(\xi, \cdot), L(\xi', \cdot)) \leq M \|\xi - \xi'\|$$

where  $d_{TV}(\nu, \nu')$  is the total variation distance between two probability measures  $\nu, \nu'$  on  $\mathbb{R}^r$ .

(L-5) There exist  $C_L \geq 0$ ,  $h_L : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  and a measurable function  $\bar{\phi}_L : \mathbb{R}^r \rightarrow \mathbb{R}_{\geq 0}$  such that the following assertions hold for  $L$ , its kernel density  $\ell$  introduced in (L-1) and its drift function introduced in (L-3).

- (i) For all  $\xi' \in \mathbb{R}^r$ , the mapping  $\xi \mapsto \ell(\xi, \xi')$  is continuous.
- (ii) We have  $\sup_{\xi, \xi' \in \mathbb{R}^r} \ell(\xi, \xi') < \infty$ .
- (iii) For all  $\xi \in \mathbb{R}^r$ , we have  $1 + \|\xi\| \leq \bar{\phi}_L(\xi)$ .
- (iv) For all  $(\xi, \xi', \xi'') \in \mathbb{R}^{3r}$ ,

$$\left| \ln \frac{\ell(\xi; \xi'')}{\ell(\xi'; \xi'')} \right| \leq h_L(\|\xi - \xi'\|) e^{C_L(1 + \|\xi\| \vee \|\xi'\|)} \bar{\phi}_L(\xi'').$$

- (v)  $h_L(u) = O(u)$  as  $u \rightarrow 0$ .
- (vi) If  $C_L = 0$ , then  $\ln^+ \bar{\phi}_L \lesssim V_L$ . Otherwise,  $\bar{\phi}_L \lesssim V_L$ .

(L-6) We have  $L(v; \{f_{1:d}(\cdot) \in A\}) < 1$  for all  $v \in \mathbb{V}$  and affine hyperplanes  $A \subset \mathbb{R}^d$ .

We can now state our main result for the PARX model.

**Theorem 7.** Consider the PARX( $p, q$ ) model of Definition 6, seen as the VLODM of Example 3. Suppose that (L-1)–(L-4) hold, and that, for all  $\theta = (\omega, a_{1:p}, b_{1:q}, \gamma_{1:d}) \in \Theta$ , we have

$$\sum_{i=1}^p a_i + \sum_{i=1}^q b_i < 1. \quad (2.22)$$

Then the following assertions hold.

(i) for all  $\theta \in \Theta$ , there exists a unique stationary solution  $\{(X_k, Y_k, \Xi_k) : k \in \mathbb{Z}_{\geq 0}\}$  to (2.1), that is, (A-1) holds for the VLODM of Example 3. Moreover, (A-2) holds with  $V_X(\bar{x}) = x + V_L(\xi)$  for all  $\bar{x} = (x, \xi) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^r$ .

(ii) Suppose moreover that (L-5) holds. Then, for any  $x_1^{(i)} \in \mathbb{R}_{\geq 0}$ ,  $\xi_1^{(i)} \in \mathbb{R}^r$  and  $y_1^{(i)} \in \mathbb{Z}_{\geq 0}$ , setting  $z^{(i)} = ((x_1^{(i)}, \xi_1^{(i)}), \dots, (x_1^{(i)}, \xi_1^{(i)}), (y_1^{(i)}, \xi_1^{(i)}), \dots, (y_1^{(i)}, \xi_1^{(i)})) \in (\mathbb{R}_{\geq 0} \times \mathbb{R}^r)^p \times (\mathbb{Z}_{\geq 0} \times \mathbb{R}^r)^{q-1}$ , the MLE  $\hat{\theta}_{z^{(i)}, n}$  as defined by (2.9) is equivalence-class consistent, that is, (2.11) holds for any  $\theta_* \in \Theta$ .

(iii) Suppose in addition that (L-6) holds. Then the MLE  $\hat{\theta}_{z^{(i)}, n}$  is consistent.

The proof is postponed to Section 4.6. Let us briefly comment on our assumptions.

**Remark 5.** Contrary to the previous example, the PARX model requires an additional Markov kernel  $L$  and therefore additional assumptions on this kernel. We briefly comment on them hereafter and compare our assumptions to those in [1].

- (i) Assumptions (L-1)–(L-3) are classical assumptions on Markov kernels to ensure its stability and ergodicity. In [1, Assumption 2], a different approach is used and instead a first order contraction of the random iterative functions defining the Markov transition is used. Their assumption would typically imply our (L-3) with  $V_L(\xi) = 1 + \|\xi\|^s$  for some  $s \geq 1$ , with their Assumption(3)(ii) implying our (L-3)(iii).
- (ii) Our Assumption (L-4) is inherited from our approach for proving ergodicity using the embedding of the PARX model into a VLODM model detailed in Example 3. It is not clear to us whether it can be omitted for proving ergodicity. This question is left for future work. Note that we provide another formulation of this assumption in Remark 10, see Section 4.6. There is no equivalent of our Assumption (L-4) in [1] as their technique for proving ergodicity is different. However, although we have the same condition (2.22) as them for ergodicity, see their Assumption 3(i), they require an additional condition, Assumption 3(iii), which we do not need here. This condition involves both the coefficients  $a_i$  and  $b_i$  and the contraction constant  $\rho$  used in their stability assumption for the Markov transition of the covariates. In contrast, our condition on the parameters  $a_i$  and  $b_i$  of the model, merely imposed through (2.22), are completely separated from our assumptions (L-1)–(L-4) on the dynamics of the covariates.
- (iii) Assumptions (L-5) and (L-6) are not required to get ergodicity. they are needed to establish the equivalence class consistency of the MLE and the

identifiability of the model. Like our Assumption (L-4) for proving ergodicity, Assumption (L-5) is inherited from the embedding of the PARX model into a VLDM model here used for proving the equivalent class consistency. Assumption (L-6) is a mild and natural identifiability assumption. It basically says that the covariates  $f_1(\xi_k), \dots, f_d(\xi_k)$  are not linearly related conditionally to  $\xi_{k-1}$ . If they were, it would suggest using a smaller set of covariates. The identifiability condition of [1] is different as it involves a condition both on the covariate distribution and on the parameters  $a_1, \dots, a_p$  and  $b_1, \dots, b_q$ , see their Assumption 5.

To conclude this section, let us carefully examine the simple case where the covariates are assumed to follow a Gaussian linear dynamic and see, in this specific case, how our assumptions compare to that of [1]. More precisely, assume that  $L$  is defined by the following equation on the exogenous variables

$$\Xi_t = \aleph \Xi_{t-1} + \sigma \eta_t, \tag{2.23}$$

where  $\aleph$  is an  $r \times r$  matrix with spectral radius  $\rho(\aleph) \in (0, 1)$ ,  $\sigma > 0$ , and  $\eta_t \sim \mathcal{N}(0, I_r)$ . Then Assumption (L-1) holds with

$$\ell(\xi, \xi') = (2\pi\sigma^2)^{-r/2} e^{-\|\xi' - \aleph\xi\|^2 / (2\sigma^2)},$$

and  $\nu_L$  being the Lebesgue measure on  $\mathbb{R}^r$ . It is straightforward to check (L-2) and (L-3) with  $V_L(\xi) = e^{\lambda \|\xi\|}$  for any  $\lambda > 0$ .

Let us now check that Assumption (L-4) holds. First note that, setting  $f(x) = e^{-x^2/2}$ , we have  $\ell(\xi, \xi'') = \frac{1}{(2\pi\sigma^2)^{r/2}} f(\|\xi'' - \aleph\xi\|)$ . Then, for all  $\xi, \xi', \xi'' \in \mathbb{R}^r$  such that  $\|\xi - \xi'\| \leq \epsilon$ , which implies  $\|\aleph(\xi - \xi')\| \leq \|\aleph\|\epsilon$ , where  $\|\aleph\|$  denotes the operator norm of  $\aleph$ , and thus

$$\begin{aligned} |\ell(\xi, \xi'') - \ell(\xi', \xi'')| &\leq \frac{\|\aleph\|}{(2\pi\sigma^2)^{r/2}} \|\xi - \xi'\| \sup_{\|\xi'' - \aleph\xi\| \leq \|\xi'' - \aleph\xi'\| + \|\aleph\|\epsilon} |f'(x)| \\ &\leq \frac{\|\aleph\|}{(2\pi\sigma^2)^{r/2}} \|\xi - \xi'\| (\|\xi''\| + \|\aleph\|\epsilon) e^{-(\|\xi'' - \aleph\xi\|)^2_+ / 2}. \end{aligned}$$

Then we get that for all  $\xi, \xi' \in \mathbb{R}^r$  such that  $\|\xi - \xi'\| \leq \epsilon$ ,

$$d_{\text{TV}}(L(\xi, \cdot), L(\xi', \cdot)) = \frac{1}{2} \int_{\mathbb{R}^r} |\ell(\xi, \xi'') - \ell(\xi', \xi'')| d\xi'' \leq M_0 \|\xi - \xi'\|,$$

for some positive constant  $M_0$  only depending on  $\|\aleph\|$ ,  $\sigma$  and  $\epsilon$ . Finally, for all  $\xi, \xi' \in \mathbb{R}^r$ ,

$$d_{\text{TV}}(L(\xi, \cdot), L(\xi', \cdot)) \leq (M_0 \mathbb{1}_{\|\xi - \xi'\| \leq \epsilon} + \epsilon^{-1} \mathbb{1}_{\|\xi - \xi'\| > \epsilon}) \|\xi - \xi'\|,$$

and Assumption (L-4) holds.

Let us now check Assumption (L-5). Assumptions (L-5)(i) and (L-5)(ii) are immediate. For all  $(\xi, \xi', \xi'') \in \mathbb{R}^{3r}$ , we have

$$\left| \ln \frac{\ell(\xi; \xi'')}{\ell(\xi'; \xi'')} \right| = \frac{1}{2\sigma^2} \left| \|\xi'' - \aleph\xi\|^2 - \|\xi'' - \aleph\xi'\|^2 \right|$$

$$\begin{aligned} &\leq \frac{\|\aleph\|}{2\sigma^2} \|\xi - \xi'\| (1 + \|\aleph\| (\|\xi\| + \|\xi'\|)) (1 + \|\xi''\|) \\ &\leq h_L(\|\xi - \xi'\|) e^{C_L(1 + \|\xi\| \vee \|\xi'\|)} \bar{\phi}_L(\xi''). \end{aligned}$$

Thus (L-5)(iv) holds with  $h_L(u) = (\|\aleph\|/2\sigma^2)u$ ,  $C_L = \|\aleph\|$  and  $\bar{\phi}_L(\xi) = 1 + \|\xi\|$ . Then, (L-5)(iii) and (L-5)(v) follow from these choices of  $h_L$  and  $\bar{\phi}_L$ . We also get (L-5)(vi) since  $V_L(\xi) = e^{\lambda\|\xi\|}$  for some  $\lambda > 0$ .

Finally (L-6) immediately holds since  $\nu_L$  is the Lebesgue measure on  $\mathbb{R}^r$ .

Having shown (L-1)–(L-6) for covariates satisfying the dynamics (2.23), Theorem 7 applies in this case under the sole condition (2.22). In comparison, checking the assumptions for ergodicity in [1] require the additional assumption

$$\sum_{i=1}^p a_i + \sum_{i=1}^q b_i < 1 - \left( \left( 1 - \sum_{i=1}^p a_i \right) (\rho(\aleph) - b_1) \right)_+,$$

see their Assumption 3(iii).

### 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 [13]. 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  $\tilde{\mathbb{P}}^{\theta_\star}$  as defined in Definition 3.
- Step 2** Establish that, as the number of observations  $n \rightarrow \infty$ , the normalized log-likelihood  $L_{z^{(i)},n}^\theta$  as defined in (2.10), for some well-chosen  $z^{(i)} \in \mathcal{X}^p$ , 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  $(\mathcal{Y}^{\mathbb{Z}}, \mathcal{Y}^{\otimes \mathbb{Z}})$ . To define  $p^\theta(\cdot|\cdot)$ , we set, for all  $y_{-\infty:0} \in \mathcal{Y}^{\mathbb{Z} \leq 0}$  and  $y \in \mathcal{Y}$ , whenever the following limit is well defined,

$$p^\theta(y | y_{-\infty:0}) = \lim_{m \rightarrow \infty} g^\theta(\psi^\theta\langle y_{-m:0} \rangle(z^{(i)}); y). \quad (3.1)$$



**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

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

it then follows that

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

**Step 4** Using an additional argument (similar to that in [37]), deduce that the MLE  $\hat{\theta}_{z^{(i)},n}$  defined by (2.9) eventually lies in any given neighborhood of the set

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

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

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

where  $\Delta$  is the metric endowing the parameter space  $\Theta$ .

**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 [15], we provided easy-to-check conditions on first order ODMs for obtaining Step 1 to Step 4. See [15, Theorem 2] for Step 1, and [15, Theorem 1] for the following steps. In [16], 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 the conditions that we use to prove ergodicity (Step 1) and, in Section 3.3, we adapt the conditions already used in [15, 16] 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 [17].

### 3.2. Ergodicity

In this section, we provide conditions that yield stationarity and ergodicity of the Markov chain  $\{(Z_k, Y_k) : k \in \mathbb{Z}_{\geq 0}\}$ , that is, we check (A-1) and (A-2). We will set  $\theta$  to be an arbitrary value in  $\Theta$  and since this is a “for all  $\theta$  (...)” condition, to save space and alleviate the notational burden, we will drop the

superscript  $\theta$  from, for example,  $G^\theta$  and  $\psi^\theta$  and respectively write  $G$  and  $\psi$ , instead.

Ergodicity of Markov chains is usually studied using  $\varphi$ -irreducibility. This approach is well known to be quite efficient when dealing with fully dominated models; see [35]. It is not at all the same picture for integer-valued observation-driven models, where other tools need to be invoked; see [23, 13, 15] 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 after they are all listed.

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

(A-3) The  $\sigma$ -fields  $\mathcal{X}$  and  $\mathcal{U}$  are Borel ones, respectively associated to  $(X, \delta_X)$  and  $(U, \delta_U)$ , both assumed to be complete and separable metric spaces.

For an LODM( $p, q$ ), the following condition, often referred to as the *invertibility* condition, see [41], is classically assumed.

(I-1) For all  $\theta \in \Theta$ , we have  $A_{1:p}(\theta) \in \mathcal{S}_p$ ,

where  $\mathcal{S}_p$  is defined in (2.17). For an ODM( $p, q$ ) with a possibly non-linear link function, (I-1) is replaced by a uniform contracting condition on the iterates of the link function, see (A-4) below. In order to write this condition in this more general case, recall that any finite  $Y$ -valued sequence  $y, \psi^\theta \langle y \rangle$  is defined by (2.8) with the recursion (2.7). Next, we may rewrite these iterates directly in terms of  $u_{0:(n-1)}$  instead of  $y_{0:(n-1)}$ . Namely, we can define

$$\tilde{\psi}^\theta \langle u_{0:(n-1)} \rangle (z) := x_n, \text{ with } x_n \text{ defined by (2.7)} \quad (3.4)$$

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  $z \in Z$  and  $y_{0:(n-1)} \in Y^n$ . Now define, for all  $n \in \mathbb{Z}_{>0}$ , the Lipschitz constant for  $\tilde{\psi}^\theta \langle u \rangle$ , uniform over  $u \in U^n$ ,

$$\text{Lip}_n^\theta = \sup \left\{ \frac{\delta_X(\tilde{\psi}^\theta \langle u \rangle (z), \tilde{\psi}^\theta \langle u \rangle (z'))}{\delta_Z(z, z')} : (z, z', u) \in Z^2 \times U^n \right\}, \quad (3.5)$$

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

$$\delta_Z(v) = \left( \max_{1 \leq k \leq p} \delta_X \circ \Pi_k^{\otimes 2}(v) \right) \vee \left( \max_{p < k < p+q} \delta_U \circ \Pi_k^{\otimes 2}(v) \right). \quad (3.6)$$

We use the following assumption on a general link function.

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

The following assumption is mainly related to the observation kernel  $G$  and relies on the metrics introduced in (A-3) and on the iterates of the link functions defined in (3.4).

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

$$\int \sup \{g(x'; y) : x' \in X, \delta_X(x', x) < \delta\} \nu(dy) < \infty. \quad (3.7)$$

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 (3.4) is continuous on  $U$  for all  $z \in Z$ .

The definitions of weak and strong Feller in (a) and (b) correspond to Feller and strong Feller of [14, Defintion 12.1.1]. Next, we consider a classical drift condition used for showing the existence of an invariant probability distribution.

(A-6) There exist measurable functions  $V_X : X \rightarrow \mathbb{R}_{\geq 0}$  and  $V_U : U \rightarrow \mathbb{R}_{\geq 0}$  such that, setting  $V_Y = V_U \circ \Upsilon$ ,  $GV_Y \lesssim V_X$ ,  $\{V_X \leq M\}$  is compact for any  $M > 0$ , and so is  $\{V_Y \leq M\}$  if  $q > 1$ , and

$$\lim_{n \rightarrow \infty} \limsup_{M \rightarrow \infty} \sup_{z \in Z} \frac{\mathbb{E}_z [V_X(X_n)]}{M + V(z)} = 0, \quad (3.8)$$

where we defined

$$V(z) = \max_{\substack{1 \leq k \leq p \\ p < \ell < p+q}} \left\{ V_X(\Pi_k(z)), \frac{V_U(\Pi_\ell(z))}{|GV_Y|_{V_X}} \right\}. \quad (3.9)$$

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

(A-7) The conditional density of  $G$  with respect to  $\nu$  satisfies, for all  $(x, y) \in X \times Y$ ,

$$g(x; y) > 0, \quad (3.10)$$

and one of the two following assertions hold.

- (a) There exists  $y_0 \in Y$  such that  $\nu(\{y_0\}) > 0$ .
- (b) The function  $(x, u) \mapsto \tilde{\psi}_u(x)$  is continuous on  $X^p \times U^q$ .

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. Under (A-8)-(i) (defined below), for any initial distribution  $\xi$  on  $(Z^2, \mathcal{Z}^{\otimes 2})$ , let  $\hat{\mathbb{E}}_\xi$  denote the expectation (operator) associated to the distribution of  $\{X_k, X'_k, U_{k'}, U'_{k'} : k > -p, k' > -q\}$  satisfying  $(X_{(-p+1):0}, U_{(-q+1):-1}, X'_{(-p+1):0}, U'_{(-q+1):-1}) \sim \xi$  and, for all  $k \in \mathbb{Z}_{\geq 0}$ ,

$$\begin{aligned} Y_k | \mathcal{F}'_k &\sim \underline{G}(X_k, X'_k; \cdot) \quad \text{and} \quad Y'_k = Y_k, \\ X_{k+1} &= \tilde{\psi}_{U_{(k-q+1):k}}(X_{(k-p+1):k}) \quad \text{and} \quad U_k = \Upsilon(Y_k), \\ X'_{k+1} &= \tilde{\psi}_{U'_{(k-q+1):k}}(X'_{(k-p+1):k}) \quad \text{and} \quad U'_k = \Upsilon(Y'_k). \end{aligned} \quad (3.11)$$

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

(A-8) There exist measurable functions  $\alpha : X^2 \rightarrow [0, 1]$ ,  $W_X : X^2 \rightarrow [1, \infty)$ ,  $W_U : U \rightarrow \mathbb{R}_{\geq 0}$  and a Markov kernel  $\underline{G}$  on  $X^2 \times \mathcal{Y}$ , dominated by  $\nu$  with kernel density  $\underline{g}(x, x'; y)$  such that, setting  $W_Y = W_U \circ \Upsilon$ , we have  $\underline{G}W_Y \lesssim W_X$  and the three following assertions hold.

(i) For all  $(x, x') \in X^2$  and  $y \in Y$ ,

$$\min \{g(x; y), g(x'; y)\} \geq \alpha(x, x') \underline{g}(x, x'; y). \quad (3.12)$$

(ii) The function  $W_X$  is symmetric on  $X^2$ ,  $W_X(x, \cdot)$  is locally bounded for all  $x \in X$ , and  $W_U$  is locally bounded on  $U$ .

(iii) We have  $1 - \alpha \leq \delta_X \times W_X$  on  $X^2$ .

And, defining, for all  $v = (z, z') \in Z^2$ ,

$$W(v) = \max \left\{ W_X \circ \Pi_k^{\otimes 2}(v), \frac{W_U(\Pi_\ell(\tilde{z}))}{|\underline{G}W_Y|_{W_X}} : \begin{array}{l} 1 \leq k \leq p \\ \tilde{z} \in \{z, z'\} \\ p < \ell < p+q \end{array} \right\}, \quad (3.13)$$

one of the two following assertions holds.

$$(iv) \lim_{\zeta \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \sup_{v \in Z^2} \frac{\hat{\mathbb{E}}_v [W_X(X_n, X'_n)]}{W^\zeta(v)} \leq 0.$$

$$(v) \lim_{n \rightarrow \infty} \limsup_{M \rightarrow \infty} \sup_{v \in Z^2} \frac{\hat{\mathbb{E}}_v [W_X(X_n, X'_n)]}{M + W(z)} = 0 \text{ and, for all } r = 1, 2, \dots, \text{ there exists } \tau \geq 1 \text{ such that } \sup_{v \in Z^2} \hat{\mathbb{E}}_v [W_X(X_r, X'_r)] / W^\tau(v) < \infty.$$

**Remark 6.** Let us comment briefly on these assumptions.

(1) In many examples, (3.12) is satisfied with

$$\underline{g}(x, x'; y) = g(\phi(x, x'); y) \quad (3.14)$$

where  $\phi$  is a measurable function from  $X^2$  to  $X$ , in which case,  $\underline{G}W_Y$  should be replaced by  $GW_Y \circ \phi$  in (A-8).

(2) If  $\tilde{\psi}^\theta$  is of the form (2.2) with  $p' = q' = 1$ , then (A-4) is equivalent to (I-1).

(3) If  $q = 1$ , the terms depending on  $\ell$  both in (3.9) and (3.13) vanish. We can take  $V_U = W_U = 0$  without loss of generality in this case.

(4) 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  $y \in Y$ . But then (3.7) gives that  $G$  is also strong Feller by dominated convergence.

(5) Note that Condition (3.7) holds 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 8.** 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  $\pi$  which makes  $\mathbb{P}_\pi$  shift invariant. Moreover it satisfies  $\mathbb{E}_\pi[V_X(X_0)] < \infty$ . 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 4.2.

The following lemma provides a general way for constructing the instrumental functions  $\alpha$  and  $\phi$  that appear in (A-8) and in Remark 6-(1). The proof can be easily adapted from [15, Lemma 1] and is thus omitted.

**Lemma 9.** Suppose that  $X = C^S$  for some measurable space  $(S, \mathcal{S})$  and  $C \subseteq \mathbb{R}$ . Thus for all  $x \in X$ , we write  $x = (x_s)_{s \in S}$ , where  $x_s \in C$  for all  $s \in S$ . Suppose moreover that for all  $x = (x_s)_{s \in S} \in X$ , 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  $s \in S$ . This means that for all  $t \in C$ ,  $y \mapsto j(t)h(t; y)$  is a density with respect to  $\nu$  and there exists a probability measure  $\mu$  on  $(S, \mathcal{S})$  such that

$$g(x; y) = \int_S j(x_s)h(x_s; y)\mu(ds), \quad y \in Y. \tag{3.15}$$

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

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

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

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

Then  $\alpha$  and  $\phi$  defined above satisfy (A-8)(i) and (3.14).

### 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 [15, Theorem 1] and [16, 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  $y \in Y$ , the function  $(\theta, x) \mapsto g^\theta(x; y)$  is continuous on  $\Theta \times X$ .
- (B-2) For all  $y \in Y$ , the function  $(\theta, z) \mapsto \psi^\theta(y)(z)$  is continuous on  $\Theta \times Z$ .

(B-3) There exist  $x_1^{(i)} \in X$ ,  $u_1^{(i)} \in U$ , a closed set  $X_1 \subseteq X$ ,  $C \geq 0$ ,  $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  and a measurable function  $\bar{\phi} : Y \rightarrow \mathbb{R}_{\geq 0}$  such that the following assertions hold with  $z^{(i)} = (x_1^{(i)}, \dots, x_1^{(i)}, u_1^{(i)}, \dots, u_1^{(i)}) \in Z$ .

(i) For all  $\theta \in \Theta$  and  $(z, y) \in Z \times Y$ ,  $\psi^\theta \langle y \rangle (z) \in X_1$ .

(ii)  $\sup_{(\theta, x, y) \in \Theta \times X_1 \times Y} g^\theta(x; y) < \infty$ .

(iii) For all  $y \in Y$  and  $\theta \in \Theta$ ,  $\delta_X(x_1^{(i)}, \psi^\theta \langle y \rangle (z^{(i)})) \vee \delta_U(\Upsilon(y), u_1^{(i)}) \leq \bar{\phi}(y)$ .

(iv) For all  $\theta \in \Theta$  and  $(x, x', y) \in X_1 \times X_1 \times Y$ ,

$$\left| \ln \frac{g^\theta(x; y)}{g^\theta(x'; y)} \right| \leq h(\delta_X(x, x')) e^{C(\delta_X(x_1^{(i)}, x) \vee \delta_X(x_1^{(i)}, x'))} \bar{\phi}(y), \quad (3.16)$$

(v)  $h(u) = O(u)$  as  $u \rightarrow 0$ .

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

**Remark 7.** If we consider a VLODM 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  $x_1^{(i)}$ , on (the compact set)  $\Theta$  and on the choice of the norm  $|\cdot|$  on  $U = \mathbb{R}^q$ .

**Remark 8.** In the case where the observations are discrete, one usually take  $\nu$  to be the counting measure on the at most countable space  $Y$ . In this case,  $g^\theta(x; y) \in [0, 1]$  for all  $\theta, x$  and  $y$  and Condition (B-3)(ii) trivially holds whatever  $X_1$  is.

We have the following result, whose proof is postponed to Section 4.3.

**Theorem 10.** 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  $\hat{\theta}_{z^{(i)}, n}$  defined by (2.9) is equivalence-class consistent, that is, the convergence (2.11) holds for any  $\theta_x \in \Theta$ .

## 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  $u \in U$  and  $y \in Y$ , we denote by  $\tilde{\Psi}_u^\theta$  and  $\tilde{\Psi}_y^\theta$  the two  $Z \rightarrow Z$  mappings defined by

$$\tilde{\Psi}_u^\theta : z = z_{1:(p+q-1)} \mapsto \begin{cases} \left( z_{2:p}, \tilde{\psi}^\theta \langle u \rangle (z), z_{(p+2):(p+q-1)}, u \right) & \text{if } q > 1 \\ \left( z_{2:p}, \tilde{\psi}^\theta \langle u \rangle (z) \right) & \text{if } q = 1, \end{cases} \quad (4.1)$$

$$\Psi_y^\theta = \tilde{\Psi}_{\Upsilon(y)}^\theta. \tag{4.2}$$

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

$$\Psi^\theta \langle y_{0:k} \rangle = \Psi_{y_k}^\theta \circ \Psi_{y_{k-1}}^\theta \circ \dots \circ \Psi_{y_0}^\theta. \tag{4.3}$$

Note in particular that  $\psi^\theta \langle y_{0:k} \rangle$  defined by (2.8) with the recursion (2.7) can be written as

$$\psi^\theta \langle y_{0:k} \rangle = \Pi_p \circ \Psi^\theta \langle y_{0:k} \rangle, \tag{4.4}$$

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

$$\Psi^\theta \langle y_{0:k} \rangle(z) = \left( (\psi^\theta \langle y_{0:j} \rangle(z))_{k-p < j \leq k}, u_{(k-q+2):k} \right), \tag{4.5}$$

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 (2.4)), Model (2.1) can be rewritten as, for all  $k \in \mathbb{Z}_{\geq 0}$ ,

$$\begin{aligned} Y_k \mid \mathcal{F}_k &\sim G^\theta(\Pi_p(Z_k); \cdot), \\ Z_{k+1} &= \Psi_{Y_k}^\theta(Z_k). \end{aligned} \tag{4.6}$$

By this representation, the  $\text{ODM}(p, q)$  is thus embedded in an  $\text{ODM}(1, 1)$ . This in principle allows us to apply the same results obtained for the class of  $\text{ODMs}(1, 1)$  in [15] to the broader class of  $\text{ODMs}(p, q)$ . Not all but some conditions written above for an  $\text{ODM}(p, q)$  indeed easily translate to the embedded  $\text{ODM}(1, 1)$ . Take for instance (A-4). By (3.5), (3.6) and (4.5), we have, for all  $n \in \mathbb{Z}_{>0}$ , using the convention  $\text{Lip}_m^\theta = 1$  for  $m \leq 0$ ,

$$\sup_{y \in \mathcal{Y}^n, z, z' \in \mathcal{Z}} \frac{\delta_{\mathcal{Z}}(\Psi^\theta \langle y \rangle(z), \Psi^\theta \langle y \rangle(z'))}{\delta_{\mathcal{Z}}(v)} \leq \mathbb{1}_{\{n < q\}} \vee \left( \max_{0 \leq j < p} \text{Lip}_{n-j}^\theta \right). \tag{4.7}$$

Hence the same assumption (A-4) will hold for the embedded  $\text{ODM}(1, 1)$ . As an  $\text{ODM}(1, 1)$ , the bivariate process  $\{(Z_k, Y_k) : k \in \mathbb{Z}_{\geq 0}\}$  is a Markov chain on the space  $(\mathcal{Z} \times \mathcal{Y}, \mathcal{Z} \otimes \mathcal{Y})$  with transition kernel  $K^\theta$  satisfying, for all  $(z, y) \in \mathcal{Z} \times \mathcal{Y}$ ,  $A \in \mathcal{Z}$  and  $B \in \mathcal{Y}$ ,

$$K^\theta((z, y); A \times B) = \int \mathbb{1}_{A \times B}(\Psi_y^\theta(z), y') G^\theta(\Pi_p(z); dy'). \tag{4.8}$$

**Remark 9.** Note that (A-1) is equivalent to saying that the transition kernel  $K^\theta$  of the complete chain admits a unique invariant probability measure  $\pi^\theta$  on  $\mathcal{Z} \times \mathcal{Y}$ . Moreover the resulting  $\pi_X^\theta$  and  $\pi_Y^\theta$  can be obtained by projecting  $\pi^\theta$  on any of its  $\mathcal{X}$  component and any of its  $\mathcal{Y}$  component, respectively.

Note also that, by itself, the process  $\{Z_k : k \in \mathbb{Z}_{\geq 0}\}$  is a Markov chain on  $(\mathcal{Z}, \mathcal{Z})$  with transition kernel  $R^\theta$  defined by setting, for all  $z \in \mathcal{Z}$  and  $A \in \mathcal{Z}$ ,

$$R^\theta(z; A) = \int \mathbb{1}_A(\Psi_y^\theta(z)) G^\theta(\Pi_p(z); dy). \tag{4.9}$$

#### 4.2. Proof of Theorem 8

The scheme of proof of this theorem is somewhat similar to that of [15, 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 [15, 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  $(Z, \delta_Z)$  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  $(\lambda, \beta) \in (0, 1) \times \mathbb{R}_{\geq 0}$  and a measurable function  $V : Z \rightarrow \mathbb{R}_{\geq 0}$  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  $z_\infty \in Z$  such that, for any  $z \in Z$  and any neighborhood  $\mathcal{N}$  of  $z_\infty$ ,  $R^m(z, \mathcal{N}) > 0$  for at least one positive integer  $m$ .
- (E-4) There exists a Markov kernel  $\bar{R}$  on  $(Z^2 \times \{0, 1\}, \mathcal{Z}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\}))$ , a Markov kernel  $\hat{R}$  on  $(Z^2, \mathcal{Z}^{\otimes 2})$ , measurable functions  $\bar{\alpha} : Z^2 \rightarrow [0, 1]$  and  $W : Z^2 \rightarrow [1, \infty)$  symmetric, and real numbers  $(D, \zeta_1, \zeta_2, \rho) \in (\mathbb{R}_+)^3 \times (0, 1)$  such that for all  $v = (z, z', u) \in X^2 \times \{0, 1\}$  and  $n \geq 1$ ,

$$1 - \bar{\alpha} \leq \delta_Z \times W \quad \text{on } Z^2, \quad (4.10)$$

$$\forall z \in Z, \exists \gamma > 0, \sup \{W(z, z') : \delta_Z(z, z') < \gamma\} < \infty, \quad (4.11)$$

$$\begin{cases} \bar{R}(v; \cdot \times Z \times \{0, 1\}) = R(z, \cdot) & \text{and} \\ \bar{R}(v; Z \times \cdot \times \{0, 1\}) = R(z', \cdot), \end{cases} \quad (4.12)$$

$$\bar{R}(v; \cdot \times \{1\}) = \bar{\alpha}(z, z') \hat{R}((z, z'), \cdot), \quad (4.13)$$

$$\hat{R}^n((z, z'); \delta_Z) \leq D \rho^n \delta_Z(z, z'), \quad (4.14)$$

$$\hat{R}^n((z, z'); \delta_Z \times W) \leq D \rho^n \delta_Z^{\zeta_1}(z, z') W^{\zeta_2}(z, z'). \quad (4.15)$$

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

*Proof.* By [45, Theorem 2], Assumption (E-2) implies that the transition kernel  $R^m$  admits an invariant probability distribution denoted hereafter by  $\pi_m$ . 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



$n \in \mathbb{Z}_{\geq 0}$ ,

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

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

**Proposition 12.** *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 [13, 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  $\alpha, Q, \bar{Q}$  and  $Q^\#$  being our  $\bar{\alpha}, R, \bar{R}$  and  $\hat{R}$ ).  $\square$

Hence it is now clear that the conclusion of Theorem 8 holds if we can apply both Lemma 11 and Proposition 12. This is done according to the following successive steps.

- Step 1** Under (A-3), the metric (3.6) makes  $(Z, \delta_Z)$  locally compact, complete and separable, hence (E-1) holds true.
- Step 2** Prove (E-2): this is done in Lemma 13 using (A-3), (A-5), (A-6) and the fact that, for all  $y \in Y$ ,  $\psi\langle y \rangle$  is continuous on  $Z$ , as a consequence of  $\text{Lip}_1 < \infty$  in (A-4).
- Step 3** Prove (E-3): this is done in Lemma 14, using (A-3), (A-4) and (A-7).
- Step 4** Define  $\bar{\alpha}$  and prove (4.10) and (4.11) in (E-4) with  $W$  as in (3.13): 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 (4.12) and (4.13) in (E-4): this is done in Lemma 15 using (A-8)(i);
- Step 6** Finally, we need to establish the additional properties of this  $\hat{R}$  required in (E-4), namely, (4.14) and (4.15). This will be done in the final part of this section using the additional Lemma 16.

Let us start with **Step 2**.

**Lemma 13.** *If for all  $y \in Y$ ,  $\psi\langle y \rangle$  is continuous on  $Z$ , then (A-5) and (A-6) imply (E-2).*

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

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

Let us define  $\tilde{f} : Z \times Y \rightarrow \mathbb{R}$  by setting, for  $y \in Y$  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, we define  $\tilde{F} : Z \times X \rightarrow \mathbb{R}$  by setting, for all  $z \in Z$  and  $x \in X$ ,

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

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

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

Since  $z \mapsto \tilde{f}(z, y)$  is continuous for all  $y$ , we have  $A(z, z', x') \rightarrow 0$  as  $(z', x') \rightarrow (z, x)$  by (3.7) and dominated convergence. We have  $B(z, x, x') \rightarrow 0$  as  $x' \rightarrow x$ , 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  $m \in \mathbb{Z}_{>0}$ .

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

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

where we used that  $V_U(U_{n-k}) = V_U(\Upsilon(Y_{n-k})) = V_Y(Y_{n-k})$  and  $\mathbb{E}_z [V_Y(Y_{n-k})] = \mathbb{E}_z [GV_Y(X_{n-k})] \leq |GV_Y|_{V_X} \mathbb{E}_z [V_X(X_{n-k})]$ , which is valid for  $n - k \geq 0$ . Now, by (3.8), for any  $\lambda \in (0, 1)$ , we can find  $m \in \mathbb{Z}_{\geq 0}$  and  $M > 0$  such that  $\mathbb{E}_z [V_X(X_{m-k})] \leq \lambda(V(z) + M)/(2(p \vee q))$  for  $0 \leq k < p \vee q$ . Hence  $R^m V \leq \lambda V + \beta$  for  $\beta = M\lambda$ .  $\square$

We now proceed with **Step 3**.

**Lemma 14.** (A-3), (A-4) and (A-7) imply (E-3).

*Proof.* We separate the proof in two cases: first we assume (A-7)(a) and secondly, we assume (A-7)(b).

**Case 1:** Assume (A-7)(a). In this case, we pick  $y_0 \in Y$  as in (A-7)(a). Set  $y_k = y_0$  and  $u_k = \Upsilon(y_0)$  for all  $k \in \mathbb{Z}_{\geq 0}$ . By Lemma 18, we have that  $\psi^\theta \langle y_{0:n} \rangle(z) = \tilde{\psi}^\theta \langle u_{0:n} \rangle(z)$  converges to the same  $x_\infty$  in  $X$ , for all  $z \in Z$ . Now set  $z_\infty = (x_\infty, \dots, x_\infty, \Upsilon(y_0), \dots, \Upsilon(y_0)) \in Z$ . Then, for any  $z \in Z$ , by (4.5) and (3.6), for any  $\delta > 0$ , there exists  $m \in \mathbb{Z}_{>0}$  such that  $\delta_Z(z_\infty, \Psi^\theta \langle y_{0:m} \rangle(z)) < \delta$  and thus

$$R^{m+1}(z; \{z' \in Z : \delta_Z(z_\infty, z') < \delta\}) \geq \mathbb{P}_z(Z_{m+1} = \Psi^\theta \langle y_{0:m} \rangle(z))$$

$$\begin{aligned} &\geq \mathbb{P}_z(Y_k = y_0, \forall k \in \{0, \dots, m\}) \\ &= \prod_{k=0}^m G^\theta(\psi^\theta\langle y_{0:k} \rangle(z); \{y_0\}) . \end{aligned}$$

By (A-7) (a) and (3.10), we have  $G^\theta(x; \{y_0\}) > 0$  for all  $x \in \mathbf{X}$ , and we conclude that  $z_\infty$  is a reachable point.

**Case 2:** Assume (A-7)(b). Since  $(\mathbf{U}, \delta_{\mathbf{U}})$  is assumed to be separable in (A-3), there exists  $u_0 \in \mathbf{U}$  such that

$$\text{for all } \delta > 0, \quad \nu(\{y \in \mathbf{Y} : \delta_{\mathbf{U}}(u_0, \Upsilon(y)) < \delta\}) > 0 . \quad (4.16)$$

For any  $u \in \mathbf{U}$  and any integers  $k \leq l$ , in the following, we denote by  $[u]_{k:l}$  the constant sequence  $u_{k:l}$  in  $\mathbf{U}^{l-k+1}$  defined by  $u_j = u$  for all  $k \leq j \leq l$ . By Lemma 18, we have that  $\tilde{\psi}^\theta\langle [u_0]_{0:n} \rangle(z)$  converges to the same  $x_\infty$  in  $\mathbf{X}$ , for all  $z \in \mathbf{Z}$ . And by (4.1), setting  $z_\infty = (x_\infty, \dots, x_\infty, u_0, \dots, u_0) \in \mathbf{Z}$ , we have that, for all  $z \in \mathbf{Z}$ ,

$$\lim_{m \rightarrow \infty} \tilde{\Psi}^\theta\langle [u_0]_{0:m} \rangle(z) = \lim_{m \rightarrow \infty} \underbrace{\tilde{\Psi}_{u_0} \circ \dots \circ \tilde{\Psi}_{u_0}}_{m \text{ times}}(z) = z_\infty \quad \text{in } \mathbf{Z} . \quad (4.17)$$

Now pick  $z \in \mathbf{Z}$  and  $\delta > 0$ . By (4.17), we can choose an integer  $m$  such that

$$\delta_{\mathbf{Z}}(z_\infty, \tilde{\Psi}^\theta\langle [u_0]_{0:m} \rangle(z)) < \delta/2 .$$

Moreover, using (A-7) (b) and (4.1), we have that, for any  $m \in \mathbb{Z}_{\geq 0}$ ,  $u_{0:m} \mapsto \tilde{\Psi}^\theta\langle u_{0:m} \rangle(z)$  is continuous on  $\mathbf{U}^{m+1}$ . Hence there exists  $\delta' > 0$  such that, for all  $y_{0:m} \in \mathbf{Y}^{m+1}$ ,  $\delta_{\mathbf{U}}(u_0, \Upsilon(y_k)) < \delta'$  for all  $0 \leq k \leq m$  implies  $\delta_{\mathbf{Z}}(\tilde{\Psi}^\theta\langle [u_0]_{0:m} \rangle(z), \tilde{\Psi}^\theta\langle \Upsilon^{\otimes(m+1)}(y_{0:m}) \rangle(z)) < \delta/2$ , which with the previous display gives that

$$\forall 0 \leq k \leq m, \quad \delta_{\mathbf{U}}(u_0, \Upsilon(y_k)) < \delta' \implies \delta_{\mathbf{Z}}(z_\infty, \tilde{\Psi}^\theta\langle \Upsilon^{\otimes(m+1)}(y_{0:m}) \rangle(z)) < \delta .$$

Thus we have

$$R^{m+1}(z; \{z : \delta_{\mathbf{Z}}(z_\infty, z) < \delta\}) \geq \mathbb{P}_z(\delta_{\mathbf{U}}(u_0, \Upsilon(Y_k)) < \delta', 0 \leq k \leq m) .$$

Applying (4.6) and then (4.16) with (3.10), we have, for all  $\ell \geq 1$ ,

$$\mathbb{P}_z(\delta_{\mathbf{U}}(u_0, \Upsilon(Y_\ell)) < \delta' | \mathcal{F}_\ell) = G^\theta(\Pi_p(Z_\ell); \{y : \delta_{\mathbf{U}}(u_0, \Upsilon(y)) < \delta'\}) > 0 .$$

It follows that for all  $\ell = m, m-1, \dots, 1$ , conditioning on  $\mathcal{F}_\ell$ ,  $\mathbb{P}_z(\delta_{\mathbf{U}}(u_0, \Upsilon(Y_k)) < \delta', 0 \leq k \leq \ell) = 0$  implies  $\mathbb{P}_z(\delta_{\mathbf{U}}(u_0, \Upsilon(Y_k)) < \delta', 0 \leq k \leq \ell-1) = 0$ . Since  $\mathbb{P}_z(\delta_{\mathbf{U}}(u_0, \Upsilon(Y_0)) < \delta') = G^\theta(\Pi_p(z); \{y : \delta_{\mathbf{U}}(u_0, \Upsilon(y)) < \delta'\}) > 0$ , we conclude that  $R^{m+1}(z; \{z : \delta_{\mathbf{Z}}(z_\infty, z) < \delta\}) > 0$  and  $z_\infty$  is a reachable point.  $\square$

We now proceed with Step 4. Let us define  $\bar{\alpha} = \alpha \circ \Pi_p^{\otimes 2}$ . By (3.6) and (3.13), we have  $W \geq W_{\mathbf{X}} \circ \Pi_p^{\otimes 2}$  and  $\delta_{\mathbf{Z}} \geq \delta_{\mathbf{X}} \circ \Pi_p^{\otimes 2}$ . Hence (4.10) follows from (A-8)(iii). Condition (4.11) directly follows from (A-8)(ii) and the definition of  $W$  in (3.13).

We now proceed with Step 5.

**Lemma 15.** Let  $\alpha : X^2 \rightarrow [0, 1]$  be a measurable function and  $\underline{G}$  be a Markov kernel on  $X^2 \times \mathcal{Y}$  satisfying (A-8)(i) and define the Markov kernel  $\bar{R}$  on  $(Z^2, \mathcal{Z}^{\otimes 2})$  by

$$\hat{R}f(v) = \int_{\mathcal{Y}} f \circ \Psi_y^{\otimes 2}(v) \underline{G}(\Pi_p^{\otimes 2}(v); dy). \quad (4.18)$$

Then one can define a Markov kernel  $\bar{R}$  on  $(Z^2 \times \{0, 1\}, \mathcal{Z}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\}))$  which satisfies (4.12) and (4.13).

*Proof.* We first define a probability kernel  $\bar{H}$  from  $Z^2$  to  $\mathcal{Y}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\})$ . Let  $(z, z') \in Z^2$  and set  $x = \Pi_p(z)$  and  $x' = \Pi_p(z')$ . We define  $\bar{H}((z, z'); \cdot)$  as the distribution of  $(Y, Y', \epsilon)$  drawn as follows. We first draw a random variable  $\bar{Y}$  taking values in  $\mathcal{Y}$  with distribution  $\underline{G}(x, x'; \cdot)$ . Then we define  $(Y, Y', \epsilon)$  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  $G(x; \cdot) = G(x'; \cdot) = \underline{G}(x, x'; \cdot)$ . In this case, we set  $(Y, Y', \epsilon) = (\bar{Y}, \bar{Y}, 1)$ .
- Suppose now that  $\alpha(x, x') < 1$ . Then, using (3.12), the functions  $(1 - \alpha(x, x'))^{-1} [g(x; \cdot) - \alpha(x, x')g(x, x'; \cdot)]$  and  $(1 - \alpha(x, x'))^{-1} [g(x'; \cdot) - \alpha(x, x')g(x, x'; \cdot)]$  are probability density functions with respect to  $\nu$  and we draw  $\Lambda$  and  $\Lambda'$  according to these two density functions, respectively. We then draw  $\epsilon$  in  $\{0, 1\}$  with mean  $\alpha(x, x')$  and, assuming  $\bar{Y}$ ,  $\Lambda$ ,  $\Lambda'$  and  $\epsilon$  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  $(z, z') \in Z^2$  and  $B \in \mathcal{Y}$ ,

$$\begin{cases} \bar{H}((z, z'); B \times \mathcal{Y} \times \{0, 1\}) = G(\Pi_p(z); B), \\ \bar{H}((z, z'); \mathcal{Y} \times B \times \{0, 1\}) = G(\Pi_p(z'); B), \end{cases} \quad (4.19)$$

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

$$\bar{R}((z, z', u); A) = \int \mathbb{1}_A(\Psi_y(z), \Psi_{y'}(z'), u_1) \bar{H}((z, z'); dy dy' du_1).$$

Then (4.19) and (4.9) immediately gives (4.12). To conclude the proof we check (4.13). We have, for all  $v = (z, z', u) \in Z^2 \times \{0, 1\}$  and  $A \in \mathcal{Z}^{\otimes 2}$ ,

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

where  $\bar{Y}$  and  $\epsilon$  are independent and distributed according to  $\underline{G}(\Pi_p^{\otimes 2}(z, z'); \cdot)$  and a Bernoulli distribution with mean  $\bar{\alpha}(z, z')$ . This, and the definition of  $\hat{R}$  in (4.18) lead to (4.13).  $\square$

In order to achieve **Step 6**, we rely on the following result which is an adaptation of [13, Lemma 9].

**Lemma 16.** *Assume that there exists  $(\varrho, D_1, \ell) \in (0, 1) \times \mathbb{R}_{\geq 0} \times \mathbb{Z}_{>0}$  such that for all  $(z, z') \in \mathbb{Z}^2$ ,*

$$\hat{R}((z, z'); \{\delta_Z \leq D_1 \delta_Z(z, z')\}) = 1, \tag{4.20}$$

$$\hat{R}^\ell((z, z'); \{\delta_Z \leq \varrho \delta_Z(z, z')\}) = 1, \tag{4.21}$$

and that  $W : \mathbb{Z}^2 \rightarrow \mathbb{R}_{\geq 0}$  satisfies

$$\lim_{\zeta \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \sup_{v \in \mathbb{Z}^2} \frac{\hat{R}^n W(v)}{W^\zeta(v)} \leq 0, \tag{4.22}$$

Then, (4.14) and (4.15) hold.

*Proof.* Note that (4.20) implies, for any non-negative measurable function  $f$  on  $\mathbb{Z}^2$ ,  $\hat{R}(\delta_Z \times f) \leq D_1 (\delta_Z \times \hat{R}f)$  and (4.21) implies  $\hat{R}^\ell(\delta_Z \times f) \leq \varrho (\delta_Z \times \hat{R}^\ell f)$ . Hence for all  $n \geq 1$ , writing  $n = k\ell + r$ , where  $r \in \{0, \dots, \ell - 1\}$  and  $k \in \mathbb{Z}_{\geq 0}$ , we get, setting  $\rho' = \varrho^{1/\ell}$ ,

$$\hat{R}^n(\delta_Z \times f) \leq D_1^r \varrho^k (\delta_Z \times \hat{R}^n f) \leq (1 \vee D_1^{\ell-1}) \varrho^{-1} \rho'^n (\delta_Z \times \hat{R}^n f).$$

Taking  $f \equiv 1$ , we get (4.14) for any  $\rho \in [\rho', 1)$ . To obtain (4.15), we take  $f = W$  and observe that (4.22) implies that, for any  $\delta > 1$ , there exists  $\zeta > 0$  such that  $\sup(\hat{R}^n W/W^\zeta) = O(\delta^n)$ . Choosing  $\delta$  small enough to make  $\rho'\delta < 1$ , we get (4.15) with  $\rho = \rho'\delta$ ,  $\zeta_1 = 1$  and  $\zeta_2 = \zeta$ .  $\square$

We finally conclude **Step 6**. By Lemma 16, it remains to check Conditions (4.20) and (4.21), and (4.22). For all  $\ell \geq 1$ , we have  $\hat{R}^\ell(v; \{\Psi\langle y_{1:\ell} \rangle^{\otimes 2}(v) : y_{1:\ell} \in \mathbb{Y}^\ell\}) = 1$ . Using (4.7), we thus get

$$\hat{R}^\ell\left(v; \delta_Z \leq \delta_Z(v) \left( \mathbb{1}_{\{\ell < q\}} \vee \max_{(\ell-p)_+ < m \leq \ell} \text{Lip}_m \right) \right) = 1,$$

and (4.20) and (4.21) both follow from (A-4).

We now check (4.22). By (4.18) (3.11) and (3.13), for all  $n \geq q$  and  $v \in \mathbb{Z}^2$ ,  $\hat{R}^n W(v)$  can be written as

$$\begin{aligned} \hat{R}^n W(v) &= \hat{\mathbb{E}}_v \left[ \max_{0 \leq k < p} W_X(X_{n-k}, X'_{n-k}) \vee \max_{1 \leq k < q} \frac{W_Y(Y_{n-k}) \vee W_Y(Y'_{n-k})}{|\underline{G}W_Y|_{W_X}} \right] \\ &\leq \sum_{0 \leq k < p \vee q} \hat{\mathbb{E}}_v \left[ W_X(X_{n-k}, X'_{n-k}) + \frac{W_Y(Y_{n-k}) + W_Y(Y'_{n-k})}{|\underline{G}W_Y|_{W_X}} \right] \\ &\leq 3 \sum_{0 \leq k < p \vee q} \hat{\mathbb{E}}_v [W_X(X_{n-k}, X'_{n-k})], \end{aligned}$$

where we used, for  $n - k \geq 0$ ,  $\hat{\mathbb{E}}_v [W_Y(Y_{n-k})] = \hat{\mathbb{E}}_v [W_Y(Y'_{n-k})] = \hat{\mathbb{E}}_v [GW_Y(X_{n-k}, X'_{n-k})]$ . We directly get that (iv) implies (4.22). As for the two conditions in (v), the first one implies that, for any  $\rho \in (0, 1)$ , there exist  $m > 0$  and  $\beta > 0$  such that  $\hat{R}^m W \leq \rho W + \beta$ , and the second one that there exist  $\tau \geq 1$  and  $C > 0$  such that  $\hat{R}^r W \leq C W^\tau$  for all  $r = 0, \dots, m - 1$ . Combining the two previous bounds, we obtain, for all  $n = mk + r$  with  $k \in \mathbb{Z}_{\geq 0}$  and  $0 \leq r < m$ ,

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

where  $C_m$  is a positive constant, not depending on  $n$ . Hence we obtain (4.22) and the proof is concluded.

#### 4.3. Proof of Theorem 10

We apply [15, 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

$$\bar{V}(x) = \max \{V_X(x_k) : 1 \leq k \leq p\}, \quad x = x_{1:p} \in \mathbb{X}^p.$$

Let us briefly check that (B-2), (B-3) and (B-4) in [15] hold. Conditions (B-2) and (B-3) correspond to our (B-1) and (B-2), noting, for the latter one, that  $\psi_y^\theta$  here corresponds to the  $\Psi_y^\theta$  defined in (4.2), inherited from the embedding. As for (B-4) in [15], 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  $Z_1 = X^{p-1} \times X_1 \times Y^{q-1}$  with the latter  $X_1$  as in (B-3). Also our condition (A-4), by Lemma 17 and (4.7) imply (B-4)(iii) for some  $\varrho \in (0, 1)$  by setting  $\bar{\psi}(z) = C \delta_Z(z^{(i)}, z)$  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,  $z^{(i)}$  has constant first  $p$  entries and constant  $q - 1$  last entries,

$$\bar{\psi}(\Psi_y^\theta(z^{(i)})) = C \left( \delta_X(x_1^{(i)}, \psi^\theta(y)(z^{(i)})) \vee \delta_U(u_1^{(i)}, \Upsilon(y)) \right).$$

The remaining conditions (vi), (vii) and (viii) of (B-2) follow directly from (3.16), (v) and (vi) in (B-3). All the conditions of [15, Theorem 1] are checked and this result gives that, for all  $\theta, \theta_* \in \Theta$ ,  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.,  $p^\theta(y|Y_{-\infty:0})$  defined as in (3.1) is well defined for all  $y \in Y$ , that, if  $\theta = \theta_*$ , it is the density of  $Y_1$  given  $Y_{-\infty:0}$  with respect to  $\nu$ , and also that the MLE  $\hat{\theta}_{z^{(i)}, n}$  satisfies (3.3) with  $\Theta^*$  defined by (3.2). Finally, by [16, Theorem 3], we also obtain that  $\Theta^* = [\theta_*]$  and the proof is concluded.

#### 4.4. Proof of Theorem 4

We prove (i), (ii) and (iii) successively.

*Proof of (i).* We apply [Theorem 8](#) with  $V_X(x) = e^{\tau|x|}$  for some arbitrary  $\tau > 0$ . Note that [Remark 2\(1\)](#) gives [\(I-1\)](#), which, by [Remark 6\(2\)](#), gives [\(A-4\)](#). [Remark 6\(5\)](#) 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  $V_X(x) = e^{\tau|x|}$ . We can further set  $V_U(u) = e^{\tau|u|}$ , hence  $V_Y(y) = e^{\tau|\ln(1+y)|}$  and, by [Lemma 19](#), we then have  $GV_Y(x) \leq 2e^{(1+x)\tau}$  so that  $|GV_Y|_{V_X} \leq 2e^\tau$ . With these definitions, [\(3.9\)](#) leads to

$$V(z) \geq (2e^\tau)^{-1} e^{\tau|z|_\infty}, \quad z \in Z, \tag{4.23}$$

where  $|z|_\infty$  denotes the max norm of  $z \in \mathbb{R}^{p+q-1}$ . Now, to bound  $\mathbb{E}_z[V_X(X_n)]$  as  $n$  grows, we see  $V_X(X_n)$  as  $e^{\tau|\lambda(Z_n)|}$  with the specific  $\lambda = \Pi_p$  and, for any linear form  $\lambda$  on  $Z$ , we look for a recursion relation applying to

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

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

$$\tilde{\lambda}_z(y_0) = \lambda(x_{(-p+2):0}, \psi\langle y_0 \rangle(z), y_{(-q+1):(-1)}, y_0). \tag{4.24}$$

Observing that  $\tilde{\lambda}_z$  is an affine function, of the form  $\tilde{\lambda}_z(y) = \vartheta_0 + \vartheta y$ , we can apply [Lemma 19](#) with  $\zeta = x_0$  (and the trivial bound  $|\vartheta_0| \vee |\vartheta_0 + \vartheta\zeta_+| \leq |\vartheta_0| \vee |\vartheta_0 + \vartheta\zeta|$ ) and obtain that

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

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

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

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

$$\mathbb{E}_z[V_X(X_n)] \leq c^n \sum_{w \in \{0,1\}^n} e^{\tau|\Pi_p \circ \widehat{\Psi}\langle w \rangle(z)|}. \tag{4.25}$$

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  $1 \leq k < n+q$ ,

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

Note that, in this recursion, we can replace  $u_{k+1-j}$  by  $w_{k+1-j}x_{k+1-j}$  for  $k + 1 - j \geq 0$ , hence  $x_{q:(n+q)}$  satisfies the recursion (2.13) and it follows that  $\Pi_p \circ \widehat{\Psi}\langle w_{0:(n+q-1)}\rangle(z)$  can be expressed as

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

Hence Condition (2.14) implies that, for all  $z \in Z$ ,

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

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

$$\lim_{n \rightarrow \infty} \sup \left\{ |z|_\infty^{-1} \left| \Pi_p \circ \widehat{\Psi}\langle w \rangle(z) \right| : w \in \{0, 1\}^n, z \in Z \setminus \{0\} \right\} = 0. \tag{4.26}$$

Hence using (4.25), we finally obtain for a positive sequence  $\{\rho_n : n \in \mathbb{Z}_{\geq 0}\}$ ,

$$\mathbb{E}_z[V_X(X_n)] \leq (2c)^n e^{\tau \rho_n |z|_\infty}, \quad \text{with} \quad \lim_{n \rightarrow \infty} \rho_n = 0,$$

which, with (4.23), leads to, for all  $z \in Z$  and  $M > 0$ ,

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

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

$$\limsup_{M \rightarrow \infty} \sup_{z \in Z} \frac{\mathbb{E}_z[V_X(X_n)]}{M + V(z)} \leq (2c)^n (2e^\tau)^{-1} e^{\tau(\rho_n - 1)C} \rightarrow 0 \text{ as } C \rightarrow \infty.$$

Using that  $\{\rho_n : n \in \mathbb{Z}_{\geq 0}\}$  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 9 with  $C = \mathbb{R} = X$  and  $S = \{1\}$ ,  $\mu$  being the Dirac mass at point 1. For all  $(x, y) \in X \times Y$ , let  $j(x) = e^{-e^x}$  and  $h(x; u) = \frac{e^{x(e^u - 1)}}{(e^u - 1)!}$ , which  $h$  satisfies (H'-1). Hence Lemma 9 gives that (A-8)(i) and (3.14) hold with

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

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

$$1 - \alpha(x, x') = 1 - e^{-|e^x - e^{x'}|} \leq |e^x - e^{x'}| \leq e^{|x| \vee |x'|} |x - x'|.$$

We thus obtain (A-8)(iii) by setting  $W_X(x, x') = e^{|x| \vee |x'|}$ , and (A-8)(ii) also follows by setting  $W_U(y) = e^{|y|}$ . Since  $W_U$  is  $V_U$  with  $\tau = 1$ , we already saw



that  $GW_Y(x) \leq 2e^{1+x}$ , hence  $|W_Y \circ \phi|_{W_X} \leq 2e$ , and (3.13) leads to, for all  $z, z'$  in  $Z$ ,

$$W(z, z') \geq (2e)^{-1} e^{|z|_\infty \vee |z'|_\infty}. \tag{4.27}$$

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

$$\begin{aligned} \hat{\mathbb{E}}_v [e^{\tau|\lambda(Z_1)| \vee |\lambda(Z'_1)|}] &\leq \hat{\mathbb{E}}_v [e^{\tau|\lambda(Z_1)|}] + \hat{\mathbb{E}}_z [e^{\tau|\lambda(Z'_1)|}] \\ &= \mathbb{E} [e^{\tau|\tilde{\lambda}_z(\ln(1+V))|}] + \mathbb{E} [e^{\tau|\tilde{\lambda}_{z'}(\ln(1+V))|}], \end{aligned}$$

where  $V \sim \mathcal{P}(e^{\phi \circ \Pi_p^{\otimes 2}(v)})$  and  $\tilde{\lambda}_z$  is defined by (4.24). By definition of  $\phi$  above, we have  $\phi \circ \Pi_p^{\otimes 2}(v) \leq \Pi_p(z), \Pi_p(z')$ . Hence Lemma 19 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  $\mathbb{E}_z [e^{\tau|\lambda(Z_1)|}]$ , that for all  $v = (z, z') \in Z^2$ ,

$$\begin{aligned} \hat{\mathbb{E}}_v [e^{\tau(|\lambda(Z_1)| \vee |\lambda(Z'_1)|)}] &\leq c \sum_{z''=z, z'} \sum_{w=0,1} e^{\tau|\lambda \circ \hat{\Psi}(w)(z'')|} \\ &\leq 2c \sum_{w=0,1} e^{\tau(|\lambda \circ \hat{\Psi}(w)(z)| \vee |\lambda \circ \hat{\Psi}(w)(z')|)}. \end{aligned} \tag{4.28}$$

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

$$\hat{\mathbb{E}}_v [W_X^\tau(X_1, X'_1)] \leq 4c e^{\tau c_0 |v|_\infty} \text{ with } |v|_\infty := |z|_\infty \vee |z'|_\infty. \tag{4.29}$$

Thus, with (4.27), 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 (4.28) and taking  $\tau = 1$  and  $\lambda = \Pi_p$ , we thus get, for all  $n \in \mathbb{Z}_{\geq 0}$  and  $v = (z, z') \in Z^2$ ,

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

Applying (4.26) and (4.27), we get that, for all  $v \in Z^2$ ,

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

where  $\{\rho_n : n \in \mathbb{Z}_{\geq 0}\}$  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  $M \rightarrow \infty$  uniformly over  $|v|_\infty \leq C$  for any  $C > 0$ , while, if  $\rho_n < 1$ , the second one tends to zero as  $|v|_\infty \rightarrow \infty$ . Hence, for  $n$  large enough, we have

$$\lim_{M \rightarrow \infty} \sup_{v \in Z^2} \frac{\hat{\mathbb{E}}_v [W_X(X_n, X'_n)]}{M + W(v)} = 0,$$

and (A-8)(v) follows, which concludes the proof.  $\square$

*Proof of (ii).* We apply [Theorem 10](#). We have already shown that [\(A-4\)](#), [\(A-1\)](#) and [\(A-2\)](#) hold in the proof of Assertion [\(i\)](#), with  $V_X(x) = e^{\tau|x|}$  for any  $\tau > 0$ . Assumptions [\(B-1\)](#) and [\(B-2\)](#) obviously hold for the log-linear Poisson GARCH model (see [Remark 7](#) for the second one). It now only remains to show that, using  $V_X$  as above, [\(B-3\)](#) is also satisfied. We set  $X_1 = X$  which trivially satisfies [\(B-3\)\(i\)](#). Condition [\(B-3\)\(ii\)](#) is also immediate (see [Remark 8](#)). 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  $(x, x', y) \in \mathbb{R}^2 \times \mathbb{Z}_{\geq 0}$ ,

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

We thus set  $h(u) = e^{x_1^{(i)}} u$ ,  $C = 1$  and  $\bar{\phi}(y) = A + B y$  for some adequate non-negative  $A$  and  $B$  so that [\(iii\)](#) (using [Remark 7](#) and  $\Upsilon(y) = \ln(1 + y) \leq y$ ), [\(iv\)](#) and [\(v\)](#) follow. Then we have  $G^\theta \bar{\phi}(x) \leq A + B e^x \lesssim V_X$ , provided that we chose  $\tau \geq 1$ . This gives [\(vi\)](#) and thus [\(B-3\)](#) holds true, which concludes the proof.  $\square$

*Proof of (iii).* We apply [[17](#), Theorem 17] in which:

- Condition [\(A-1\)](#) corresponds to our Assumption [\(A-1\)](#) shown in Point [\(i\)](#) above;
- Condition [\(L-3\)](#) corresponds to checking that

$$\int \ln^+(\ln(1 + y)) \pi_Y^\theta(dy) < \infty. \quad (4.30)$$

We already checked that  $G^\theta V_Y \lesssim V_X$  with  $V_Y(y) = (1 + y)^\tau$ , implying  $\int (1 + y)^\tau \pi_Y^\theta(dy) < \infty$ . Hence [\(4.30\)](#) holds.

- Condition [\(SL-1\)](#) is the same as our Condition [\(I-1\)](#), which were already proved for showing Point [\(i\)](#) above;
- Conditions [\(SL-2\)](#) and [\(SL-3\)](#) are immediately checked for the log-linear Poisson GARCH model;
- Condition [\(SL-4\)](#) is directly assumed in [\(iii\)](#);

Thus [[17](#), Theorem 17] gives that  $[\theta_\star]$  reduced to  $\{\theta_\star\}$  and Assertion [\(iii\)](#) follows from [\(ii\)](#).  $\square$

#### 4.5. Proof of [Theorem 5](#)

We prove [\(i\)](#), [\(ii\)](#) and [\(iii\)](#) successively. We take  $\nu$  is the counting measure on  $\mathbb{Z}_{\geq 0}$  so that we have

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

*Proof of (i).* As for the Log-linear Poisson GARCH( $p, q$ ), we apply [Theorem 8](#) this time with  $V_X(x) = x$ . 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 6(2). Also as for the log-linear Poisson Garch( $p, q$ ) case, Remark 6(5) 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  $\Upsilon$  is the identity mapping.

We start with (A-6), with  $V_X(x) = x$ . We can further set  $V_Y(y) = y$  since, we then have  $GV_Y(x) = rV_X(x)$  so that  $|GV_Y|_{V_X} = r$ . With these definitions, (3.9) leads to

$$V(z) \geq (1+r)^{-1} |z|_\infty, \quad z \in Z. \tag{4.32}$$

On the other hand, we have that, for all  $z = (x_{(-p+1):0}, u_{(-q+1):(-1)}) \in Z = \mathbb{R}_{\geq 0}^{p+q-1}$  and all  $n = 1, 2, \dots$ ,

$$\mathbb{E}_z[V_X(X_n)] = \mathbb{E}_z[X_n] = \omega + \sum_{k=1}^p a_k \mathbb{E}_z[X_{n-k}] + \sum_{k=1}^q b_k \mathbb{E}_z[Y_{n-k}],$$

where, in the right-hand side of this equation, if  $n - k \leq 0$ ,  $\mathbb{E}_z[X_{n-k}]$  should be replaced by  $x_{n-k}$  and if  $n - k \leq -1$ ,  $\mathbb{E}_z[Y_{n-k}]$  should be replaced by  $u_{n-k}$ . By definition of  $G$ ,  $\mathbb{E}_z[Y_{n-k}] = r\mathbb{E}_z[X_{n-k}]$ . Hence, denoting  $x_n = \mathbb{E}_z[X_n]$  we get that the sequence  $\{x_k : k \in \mathbb{Z}_{\geq 0}\}$  satisfies the recursion

$$x_n = \omega + \sum_{k=1}^{p \vee q} c_k x_{n-k}, \quad n > p \vee q,$$

where here  $c_k = a_k + r b_k \geq 0$  for  $k = 1, \dots, p \wedge q$ ,  $c_k = a_k$  if  $q < k \leq p$  and  $c_k = r b_k$  if  $p < k \leq q$ . Moreover we can clearly find a constant  $C$  only depending on  $\theta$  such that  $|x_{1:(p \vee q)}|_\infty \leq C(1 + |z|_\infty)$ . Then applying Lemma 20 and (4.32), we get that there exists  $C' > 0$  and  $\rho \in (0, 1)$  such that, for all  $n \geq 1$ ,

$$\mathbb{E}_z[X_{n+1}] \leq C' \rho^n (1 + V(z)).$$

Condition (A-6) follows.

We conclude with the proof of (A-8). Let us apply Lemma 9 with  $\mathbf{C} = (0, \infty) = \mathbf{X}$  and  $\mathbf{S} = \{1\}$ ,  $\mu$  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  $\alpha$  and  $\phi$  satisfying (A-8)(i) and (3.14) by setting, for all  $x, x' \in \mathbb{R}_{\geq 0}$

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

Next, for all  $(x, x') \in Z^2$ , 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  $W_X(x, x') = (1 \vee r)$ . All the other conditions of (A-8) are trivially satisfied in this case, taking  $W_U \equiv 1$ .  $\square$

*Proof of (ii).* We apply [Theorem 10](#). We have already shown that [\(A-4\)](#), [\(A-1\)](#) and [\(A-2\)](#) hold in the proof of [Assertion \(i\)](#), with  $V_X(x) = x$ . Assumption [\(B-1\)](#) follows from [\(4.31\)](#), and [\(B-2\)](#) from [Remark 7](#). It now only remains to show that, using  $V_X$  as above, [\(B-3\)](#) is also satisfied.

Since  $\Theta$  is compact, we can find  $\underline{\omega} > 0$  and  $\bar{r} \in \mathbb{Z}_{\geq 0}$  such that  $\omega \geq \underline{\omega}$  and  $r \leq \bar{r}$  for all  $\theta = (\omega, a_{1:p}, b_{1:q}, r) \in \Theta$ . We set  $X_1 = [\underline{\omega}, \infty)$  which then satisfies [\(B-3\)\(i\)](#). Condition [\(B-3\)\(ii\)](#) follows from [Remark 8](#).

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  $\theta \in \Theta$ ,  $(x, x') \in X_1$  and  $y \in \mathbb{Z}_{\geq 0}$ , we have

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

We set  $C = 0$ ,  $h(s) = s$  and  $\bar{\phi}(y) = A + By$  for some adequate non-negative  $A$  and  $B$  so that [\(iii\)](#) (using [Remark 7](#) and  $\Upsilon(y) = y$ ), [\(iv\)](#) and [\(v\)](#) follow. Then we have  $G^\theta \ln^+ \bar{\phi}(x) \leq A + Brx \lesssim V_X$ . This gives [\(vi\)](#) and thus [\(B-3\)](#) holds true, which concludes the proof.  $\square$

*Proof of (iii).* The proof of this point is similar to the proof of [Theorem 4\(iii\)](#) in [Section 4.4](#), except that Condition [\(4.30\)](#) is replaced by

$$\int \ln^+(|y|) \pi_V^\theta(dy) < \infty. \quad (4.33)$$

We already checked that  $G^\theta V_Y \lesssim V_X$  with  $V_Y(y) = y$ , implying  $\int y \pi_V^\theta(dy) < \infty$ . Hence [\(4.33\)](#) holds and the proof is concluded.  $\square$

#### 4.6. Proof of [Theorem 7](#)

The following remark about Assumption [\(L-4\)](#) will be useful.

**Remark 10.** If the Markov kernel  $L$  admits a kernel density  $\ell$  with respect to some measure  $\nu_L$  on Borel sets of  $\mathbb{R}^r$  as in [\(L-1\)](#), then

$$\begin{aligned} d_{\text{TV}}(L(\xi, \cdot), L(\xi', \cdot)) &= \frac{1}{2} \int |\ell(\xi; \xi'') - \ell(\xi'; \xi'')| \nu_L(d\xi'') \\ &= 1 - \int (\ell(\xi; \xi'') \wedge \ell(\xi'; \xi'')) \nu_L(d\xi''). \end{aligned}$$

It follows that Assumption [\(L-4\)](#) is equivalent to assuming that there exist a measurable function  $\alpha_L : \mathbb{R}^{2r} \rightarrow [0, 1]$  and a Markov kernel  $\underline{L}$  on  $(\mathbb{R}^r)^2 \times \mathcal{B}(\mathbb{R}^r)$  dominated by  $\nu_L$  with kernel density  $\underline{\ell}(\xi, \xi'; \xi'')$  such that

(a) for all  $(\xi, \xi', \xi'') \in \mathbb{R}^{3r}$ ,

$$\ell(\xi; \xi'') \wedge \ell(\xi'; \xi'') \geq \alpha_L(\xi, \xi') \underline{\ell}(\xi, \xi'; \xi'').$$

(b) there exists a constant  $M \geq 1$  such that for all  $(\xi, \xi') \in \mathbb{R}^{2r}$ ,

$$1 - \alpha_L(\xi, \xi') \leq M \|\xi - \xi'\|.$$

Obviously these two conditions imply (L-4). To see the converse implication, take  $\alpha_L(\xi, \xi') = \int_{\mathbb{R}^r} (\ell(\xi; \xi'') \wedge \ell(\xi'; \xi'')) \nu_L(d\xi'')$  and  $\underline{\ell}(\xi, \xi'; \xi'') = (\ell(\xi; \xi'') \wedge \ell(\xi'; \xi'')) / \alpha_L(\xi, \xi')$ .

*Proof of (i).* As explained in Example 3, the PARX model can be cast into a VLODM( $p, q$ ) and, to prove ergodicity, we can thus apply Theorem 8. Hence we now need to check Conditions (A-3), (A-4), (A-5), (A-6), (A-7) and (A-8). (A-3) always hold for a VLODM. (A-4) follows from (2.22) by Assertion (i) in Lemma 21.

Let us check (A-5), by first verifying (3.7). Using the definition of the kernel  $G$  (we drop  $\theta$  from the notation  $G^\theta$  since it does not depend on  $\theta$ ) in Example 3 b), and using (L-1),  $G$  admits a kernel density  $g$  wrt  $\nu := \nu_{\mathbb{Z}_{\geq 0}} \otimes \nu_L$ , with  $\nu_{\mathbb{Z}_{\geq 0}}$  denoting the counting measure on  $\mathbb{Z}_{\geq 0}$ , given by

$$g(\bar{x}; \bar{y}) = e^{-x} \frac{x^y}{y!} \ell(\xi; \xi'), \quad (\bar{x}, \bar{y}) = ((x, \xi), (y, \xi')) \in \bar{X} \times \bar{Y}. \tag{4.34}$$

We set, for  $(\bar{x}, \bar{x}') = ((x, \xi), (x', \xi')) \in \bar{X}^2$

$$\delta_X(\bar{x}, \bar{x}') = |x - x'| + \|\xi - \xi'\|. \tag{4.35}$$

According to assumption (L-1)(ii), there exists  $\delta > 0$  such that  $\int_{\mathbb{R}^r} \sup_{\|\xi' - \xi\| \leq \delta} h(\xi'; \xi'') \nu_L(d\xi'') < \infty$ . Thus, using the expression of  $g$  given in (4.34),

$$\begin{aligned} & \sum_{y=0}^{\infty} \int_{\mathbb{R}^r} \sup \{g((x', \xi'); (y, \xi'')), \delta_X((x', \xi'), (x, \xi)) \leq \delta\} \nu_L(d\xi'') \\ & \leq \sum_{y=0}^{\infty} e^{-x+\delta} \frac{(x+\delta)^y}{y!} \int_{\mathbb{R}^r} \sup_{\|\xi' - \xi\| \leq \delta} \{h(\xi'; \xi'')\} \nu_L(d\xi''), \\ & \leq e^{2\delta} \int_{\mathbb{R}^r} \sup_{\|\xi' - \xi\| \leq \delta} \{h(\xi'; \xi'')\} \nu_L(d\xi'') < \infty, \end{aligned}$$

which shows (3.7). Since  $L$  is weak-Feller by (L-2), (A-5)-(b) holds, and finally (A-5) is satisfied.

We now turn to checking (A-6). Define

$$\begin{aligned} V_Y(y, \xi) &= V_U \circ \Upsilon(y, \xi) = V_U(y, f_1(\xi), \dots, f_d(\xi), \xi) = y + V_L(\xi), \\ V_X(x, \xi) &= x + V_L(\xi). \end{aligned}$$

Then, using (L-3)-(iv) with  $n = 1$  and  $h = V_L$ ,

$$\frac{GV_Y}{V_X}(x, \xi) = \frac{x + LV_L(\xi)}{x + V_L(\xi)} \leq \frac{x + C_\varrho V_L(\xi) + \pi(V_L)}{x + V_L(\xi)} \leq 1 \vee (C_\varrho) + \pi(V_L),$$

so that  $GV_Y \lesssim V_X$ . Moreover,  $\{V_X \leq M\}$  is compact for any  $M > 0$ , and so is  $\{V_Y \leq M\}$  since  $\{V_L \leq M\}$  is compact for any  $M > 0$  by (L-3)-(ii). To complete the proof of (A-6), it only remains to check (3.8). Recall that in this model,  $\bar{U}_t = (Y_t, f_1(\Xi_t), \dots, f_d(\Xi_t), \Xi_t)$ . Noting that for all  $z \in Z = \bar{X}^p \times \bar{U}^{q-1}$ ,  $\mathbb{E}_z[Y_t] = \mathbb{E}_z[X_t]$  (see (2.21)), we have, using (L-3),

$$\begin{aligned} \mathbb{E}_z[X_n] &= \mathbb{E}_z \left[ \omega + \sum_{i=1}^p a_i X_{n-i} + \sum_{i=1}^q b_i Y_{n-i} \right] + L^n(\mathbf{f}_\gamma)(\Xi_{-1}) \\ &\leq \omega + \sum_{i=1}^p a_i \mathbb{E}_z[X_{n-i}] + \sum_{i=1}^q b_i \mathbb{E}_z[X_{n-i}] \\ &\quad + C\varrho^n |\mathbf{f}_\gamma|_{V_L} V_L(\Xi_{-1}) + \pi_L(\mathbf{f}_\gamma). \end{aligned}$$

Defining  $c_i$  as in Assertion (ii) of Lemma 21, we write this inequality as

$$\mathbb{E}_z[X_n] \leq \omega + \sum_{i=1}^{p \vee q} c_i \mathbb{E}_z[X_{n-i}] + C\varrho^n |\mathbf{f}_\gamma|_{V_L} V_U(\Pi_{p+q}(z)) + \pi_L(\mathbf{f}_\gamma). \quad (4.36)$$

Similarly, we have

$$\begin{aligned} \mathbb{E}_z[V_X(X_n, \Xi_{n-1})] &= \mathbb{E}_z[X_n] + L^n(V_L)(\Xi_{-1}) \\ &\leq \mathbb{E}_z[X_n] + C\varrho^n V_U(\Pi_{p+q}(z)) + \pi_L(V_L). \end{aligned} \quad (4.37)$$

Combining (4.36) and (4.37) yields that, for any  $\epsilon > 0$ , we can find  $M$  large enough such that, with  $V$  defined by (3.9), and setting

$$u_n := \sup_{z \in Z} \frac{\mathbb{E}_z[V_X(X_n, \Xi_{n-1})]}{M + V(z)}, \quad n \in \mathbb{Z}_{\geq 0},$$

we have, for all  $n \in \mathbb{Z}_{>0}$ ,  $u_n \in \mathbb{R}_{\geq 0}$  and

$$u_n \leq \sum_{i=1}^{p \vee q} c_i u_{n-i} + C\varrho^n |GV_Y|_{V_X} + \frac{1}{2}\epsilon.$$

Thus, since  $\varrho < 1$ , for  $n$  large enough,

$$u_n \leq \sum_{i=1}^{p \vee q} c_i u_{n-i} + \epsilon.$$

It follows that, for any  $\epsilon > 0$ ,

$$\limsup_{n \rightarrow \infty} \limsup_{M \rightarrow \infty} \sup_{z \in Z} \frac{\mathbb{E}_z[V_X(X_n, \Xi_{n-1})]}{M + V(z)} \leq \limsup_{n \rightarrow \infty} u_n \leq \frac{\epsilon}{1 - \sum_{i=1}^{p \vee q} c_i},$$

where, in the second inequality, we used Lemma 20 with Assertion (ii) of Lemma 21. We thus get (3.8).

Let us now check (A-7). From (L-1) (i) and (4.34), we have (3.10). Since (A-7)-(b) always holds for a VLODM, we obtain (A-7).

To complete the proof, it remains to check (A-8). In what follows, we set for  $(\bar{x}, \bar{x}') = ((x, \xi), (x', \xi')) \in \bar{X}^2$  and  $\bar{y} = (y, \xi'') \in \bar{Y}$ ,

$$\alpha(\bar{x}, \bar{x}') = e^{-|x-x'|} \alpha_L(\xi, \xi'), \quad \underline{g}(\bar{x}, \bar{x}'; \bar{y}) = e^{-x \wedge x'} \frac{(x \wedge x')^y}{y!} \underline{\ell}(\xi, \xi'; \xi''), \quad (4.38)$$

where  $\alpha_L$  is given by Remark 10 and take for  $W_X : \bar{X}^2 \rightarrow [1, \infty)$  and  $W_U : \bar{U} \rightarrow \mathbb{R}_{\geq 0}$  the constant functions equal to  $M$ , where  $M$  is the constant that appears in Assumption (L-4) and Remark 10(b).

Using (4.34), we have, for all  $(\bar{x}, \bar{x}') = ((x, \xi), (x', \xi')) \in \bar{X}^2$  and  $\bar{y} = (y, \xi'') \in \bar{Y}$ ,

$$\begin{aligned} \min\{g(\bar{x}; \bar{y}), g(\bar{x}'; \bar{y})\} &\geq e^{-x \vee x'} \frac{(x \wedge x')^y}{y!} h(\xi; \xi'') \wedge h(\xi'; \xi'') \\ &\geq e^{-|x-x'|} e^{-x \wedge x'} \frac{(x \wedge x')^y}{y!} \alpha_L(\xi, \xi') \underline{\ell}(\xi, \xi'; \xi'') \\ &\geq \alpha(\bar{x}, \bar{x}') \underline{g}(\bar{x}, \bar{x}'; \bar{y}), \end{aligned}$$

where  $\alpha$  and  $\underline{g}$  are defined in (4.38). Then (A-8)-(i) holds. Moreover, since  $W_X$  and  $W_U$  are constants, (A-8)-(ii) trivially holds. We now turn to (A-8)-(iii). By (4.38), for all  $(\bar{x}, \bar{x}') = ((x, \xi), (x', \xi')) \in \bar{X}^2$ ,

$$\begin{aligned} 1 - \alpha(\bar{x}, \bar{x}') &= 1 - e^{-|x-x'|} \alpha_L(\xi, \xi') \leq 1 + e^{-|x-x'|} (M \|\xi - \xi'\| - 1) \\ &\leq 1 - e^{-|x-x'|} + e^{-|x-x'|} M \|\xi - \xi'\| \leq M(|x - x'| + \|\xi - \xi'\|), \end{aligned} \quad (4.39)$$

where in the last inequality we have used  $1 - e^{-|x-x'|} \leq |x - x'| \leq M|x - x'|$ . Plugging the definition of  $\delta_X$  given in (4.35) into (4.39) shows (A-8)-(iii). Since  $W_X$  and  $W_U$  are constants, we obtain that  $W$  defined in (A-8) is also constant and therefore any of the two conditions (A-8)-(iv) or (A-8)-(v) trivially holds.

We have thus checked all the assumptions of Theorem 8 for the VLODM of Example 3 and this concludes the proof of Assertion (i).  $\square$

*Proof of (ii).* We apply Theorem 10. We have already shown that (A-4), (A-1) and (A-2) hold in the proof of Assertion (i), with  $V_X(x, \xi) = x + V_L(\xi)$ . Using (4.34) and the assumption (L-5)(i), we get (B-1) (here  $g$  does not depend on  $\theta$ ). Assumption (B-2) always holds for a VLODM.

To conclude the proof of (ii), it thus remains to check (B-3). Since  $\Theta$  is compact, we can find  $\underline{\omega} > 0$  such that  $\omega \geq \underline{\omega}$  for all  $\theta = (\omega, a_{1:p}, b_{1:q}, \gamma_{1:d}) \in \Theta$ . Then we set  $\bar{X}_1 = [\underline{\omega}, \infty) \times \mathbb{R}^r$  which is a closed subset of  $\bar{X}$  and for which (B-3)(i) holds (with  $Y$  replaced by  $\bar{Y}$  and  $X_1$  by  $\bar{X}_1$  since we deal with Example 3). (B-3)(ii) holds as an immediate consequence of (4.34) and (L-5)(ii). In the VLODM case, to meet (B-3)(iii), it suffices to have

$$\bar{\phi}(y, \xi) \geq c_0(1 + y + \|\xi\|), \quad y \in \mathbb{Z}_{\geq 0}, \xi \in \mathbb{R}^r,$$

where  $c_0$  is a positive constant only depending on  $z^{(i)}$ . In the following, we choose

$$\bar{\phi}(y, \xi) := (1 \vee c_0) \left( 1 + (1 \vee \underline{\omega}^{-1}) y + \bar{\phi}_L(\xi) \right), \quad y \in \mathbb{Z}_{\geq 0}, \xi \in \mathbb{R}^r, \quad (4.40)$$

which, with **(L-5)(iii)**, guaranties that the previous inequality holds, and thus **(B-3)(iii)** is verified.

Let  $\bar{x}, \bar{x}' \in \bar{X}_1$  and  $\bar{y} \in \bar{Y}$ , that is  $\bar{x} = (x, \xi)$  and  $\bar{x}' = (x', \xi')$  with  $x, x' \in [\underline{\omega}, \infty)$  and  $\xi, \xi' \in \mathbb{R}^r$ , and  $\bar{y} = (y, \xi'')$  with  $y \in \mathbb{Z}_{\geq 0}$  and  $\xi'' \in \mathbb{R}^r$ , we have, by **(4.34)**,

$$\left| \ln \frac{g(\bar{x}; \bar{y})}{g(\bar{x}'; \bar{y})} \right| \leq |x - x'| + y \left| \ln \frac{x}{x'} \right| + \left| \ln \frac{\ell(\xi; \xi'')}{\ell(\xi'; \xi'')} \right|$$

By **(L-5)(iv)** and using that  $|\ln(x/x')| \leq \underline{\omega}^{-1}|x - x'|$ , we obtain that

$$\left| \ln \frac{g(\bar{x}; \bar{y})}{g(\bar{x}'; \bar{y})} \right| \leq |x - x'| (1 + \underline{\omega}^{-1}y) + h_L(\|\xi - \xi'\|) e^{C_L(1 + \|\xi\| \vee \|\xi'\|)} \bar{\phi}_L(\xi'').$$

Thus **(B-3)(iv)** holds with  $\bar{\phi}$  defined as in **(4.40)** and for some well chosen  $c_1 > 0$  only depending on  $z^{(i)}$

$$h(u) = c_1 (u + h_L(u)) \quad \text{and} \quad C = C_L \quad (4.41)$$

Obviously **(B-3)(v)** is then satisfied. We now observe that, with  $G$  given by **Example 3 b)** and  $\bar{\phi}$  given by **(4.40)**, for all  $x \in \mathbb{R}_{\geq 0}$  and  $\xi \in \mathbb{R}^r$ , we have

$$\begin{aligned} G[\ln^+ \bar{\phi}](x, \xi) &\leq \ln(1 \vee c_0) + (1 \vee \underline{\omega}^{-1}) x + L[\ln^+ \bar{\phi}_L](\xi), \\ G\bar{\phi}(x, \xi) &\leq (1 \vee c_0) \left( 1 + (1 \vee \underline{\omega}^{-1}) x + L\bar{\phi}_L(\xi) \right), \end{aligned}$$

where, in the first inequality, we used that  $\ln^+(1 + (1 \vee \underline{\omega}^{-1}) y + \bar{\phi}_L(\xi)) \leq (1 \vee \underline{\omega}^{-1}) y + \ln^+(\bar{\phi}_L(\xi))$ . By **(L-3)(iv)** we have  $LV_L \lesssim V_L$  and thus **(L-5)(vi)** implies  $L[\ln^+ \bar{\phi}_L] \lesssim LV_L \lesssim V_L$  if  $C_L = 0$  and  $L\bar{\phi}_L \lesssim LV_L \lesssim V_L$  otherwise. The previous display with **(L-5)(vi)** yields **(B-3)(vi)** with  $C = C_L$ . (Recall that  $V_X(x, \xi) = x + V_L(\xi)$ ).  $\square$

*Proof of (iii).* To obtain **(iii)** from Assertion **(ii)** we only need to have that  $\{\theta_\star\}$  reduced to the singleton  $\{\theta_\star\}$  for all  $\theta_\star \in \Theta$ . To prove this, we use [17, Section 5.5] (or [18, Theorem 18]), where it is proved under Assumption **(L-6)** and other assumptions named **(A'-1)**, **(L-1)** and **(L-3)** which, in our setting here, respectively correspond to **(A-1)**, **(l-1)** and to having, for all  $\theta \in \Theta$ ,

$$\mathbb{E}^\theta \left[ \ln^+ \left( Y_0 + \sum_{k=1}^d f_k(\Xi_0) + \|\Xi_0\| \right) \right] < \infty. \quad (4.42)$$

We have already shown in Assertion **(i)** that **(A-1)** and **(l-1)** hold, and also **(A-2)** with  $V_X(x, \xi) = x + V_L(\xi)$ , which yield that, for all  $\theta \in \Theta$ ,  $\mathbb{E}^\theta [X_0 + V_L(\Xi_{-1})] < \infty$ , and thus

$$\mathbb{E}^\theta [Y_0 + V_L(\Xi_0)] < \infty. \quad (4.43)$$



Using that, for all  $y \in \mathbb{Z}_{\geq 0}$  and  $\xi \in \mathbb{R}^r$ ,

$$\begin{aligned} \ln^+ \left( y + \sum_{k=1}^d f_k(\xi) + \|\xi\| \right) &\leq \ln \left( 1 + \|\xi\| + \sum_{k=1}^d f_k(\xi) + y \right) \\ &\leq \sum_{k=1}^d f_k(\xi) + y + \ln(1 + \|\xi\|) , \end{aligned}$$

we get that (4.42) follows from (4.43), (L-3)(iii) and the fact that  $\ln(1 + \|\cdot\|) \lesssim \ln^+ \bar{\phi}_L \lesssim V_L$  as a consequence of (L-5)(iii) and (vi), respectively. This concludes the proof of Assertion (iii) of Theorem 7.  $\square$

### 5. Useful lemmas

The following result is used in the proof of Theorem 10. It is proven in [18, Lemma 19].

**Lemma 17.** (A-4) implies that for all  $\theta \in \Theta$ , there exist  $C > 0$  and  $\rho \in (0, 1)$  such that  $\text{Lip}_n^\theta \leq C \rho^n$  for all  $n \in \mathbb{Z}_{>0}$ .

A byproduct of Lemma 17 is the following result, which is used in the proof of Lemma 14.

**Lemma 18.** Suppose that (A-3) and (A-4) hold. Let  $\theta \in \Theta$  and  $u_0 \in \mathbb{U}$  and set  $u_k = u_0$  for all  $k \in \mathbb{Z}_{\geq 0}$ . Then there exists  $x_\infty \in \mathbb{X}$  such that for all  $z \in \mathbb{Z}$ ,  $\tilde{\psi}^\theta \langle u_{0:n} \rangle(z)$  converges to  $x_\infty$  as  $n \rightarrow \infty$  in  $(\mathbb{X}, \delta_X)$ .

*Proof.* Let  $z, z'$  in  $\mathbb{Z}$ . Denote, for all  $n \in \mathbb{Z}_{>0}$ ,  $x_n = \tilde{\psi}^\theta \langle u_{0:n} \rangle(z)$  and  $\tilde{x}_n = \psi^\theta \langle u_{0:n} \rangle(z')$ . Then, for all  $n \in \mathbb{Z}_{>0}$ ,

$$\delta_X(x_n, \tilde{x}_n) \leq \text{Lip}_n^\theta \delta_Z(z, z') , \tag{5.1}$$

$$\delta_X(x_{n+1}, x_n) \leq \text{Lip}_n^\theta \delta_Z \left( z, \tilde{\Psi}_{u_0}^\theta(z) \right) , \tag{5.2}$$

where  $\tilde{\Psi}$  is defined in (4.1). By Lemma 17, the right-hand side of (5.2) is decreasing geometrically fast and  $x_n$  converges to a point  $\psi_\infty(y_0)$  which does not depend on  $z$  by (5.1).  $\square$

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

**Lemma 19.** Let  $\vartheta \in \mathbb{R}$ . Then, for all  $\vartheta_0 \in \mathbb{R}$  and  $\zeta \in \mathbb{R}$ , if  $U \sim \mathcal{P}(e^\zeta)$ , then

$$\mathbb{E}[(1 + U)^\vartheta] \leq e^{(1+\zeta_+)\vartheta_+} \tag{5.3}$$

$$\mathbb{E}[e^{|\vartheta_0 + \vartheta \ln(1+U)|}] \leq 2e^{\vartheta_+} e^{|\vartheta_0| \vee |\vartheta_0 + \vartheta \zeta_+|} \tag{5.4}$$

$$\leq 2e^{\vartheta_+} e^{|\vartheta_0| \vee |\vartheta_0 + \vartheta \zeta'_+|} \quad \text{for all } \zeta'_+ \geq \zeta . \tag{5.5}$$

*Proof.* We separate the proof of (5.3) in three different cases by specifying the bound (5.3) 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  $\mathbb{E}[(1+U)^\vartheta] \leq e^\vartheta$ .

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

The bound in [Case 1](#) is obvious. The bound in [Case 2](#) follows from

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

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

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

Hence we get [\(5.3\)](#).

Let us now prove [\(5.4\)](#). Observe that

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

Then using [\(5.3\)](#), we get

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

We conclude [\(5.4\)](#) and [\(5.5\)](#) by observing that, for all  $a, b \in \mathbb{R}$ ,  $(a+b_+) \vee (-a+b_-) = a \vee (a+b) \vee (-a) \vee (-a-b) = |a| \vee |a+b|$ .  $\square$

The two following lemmas are straightforward and their proofs are thus omitted. They are used in the proofs of [Theorem 5](#) and [Theorem 7](#), respectively.

**Lemma 20.** *Let  $r$  be a positive integer and  $(\omega, c_{1:r}) \in \mathbb{R}^{1+r}$ . Let  $\{x_k : k \in \mathbb{Z}_{\geq 0}\}$  be a sequence satisfying*

$$x_n = \omega + \sum_{k=1}^r c_k x_{n-k}, \quad n \geq r,$$

*Suppose that the polynomial  $P(z) = 1 - \sum_{k=1}^r c_k z^k$  has no roots in  $\{z \in \mathbb{C} : |z| \leq 1\}$ . Then there exist  $\rho < 1$  and  $C > 0$  such that, for all  $n \in \mathbb{Z}_{\geq 0}$ ,*

$$|x_n - \omega/P(1)| \leq C \rho^n (1 + \max(|x_0|, \dots, |x_{r-1}|)).$$

**Lemma 21.** *For any  $a_1, \dots, a_p, b_1, \dots, b_q \in \mathbb{R}_{\geq 0}$ , Condition [\(2.22\)](#) implies the first and is equivalent to the second following assertions.*

- (i) *The polynomial  $z \mapsto 1 - \sum_{i=1}^p a_i z^i$  has no roots in  $\{z \in \mathbb{C} : |z| \leq 1\}$ .*
- (ii) *The polynomial  $z \mapsto 1 - \sum_{i=1}^{p+q} c_i z^i$  has no roots in  $\{z \in \mathbb{C} : |z| \leq 1\}$  where  $c_i = a_i \mathbb{1}(i \leq p) + b_i \mathbb{1}(i \leq q) \in \mathbb{R}_{\geq 0}$ .*

## Acknowledgements

The authors are thankful to the anonymous referee and to the editor-in-chief for insightful comments and helpful suggestions that led to improving the quality of the paper.

## References

- [1] AGOSTO, A., CAVALIERE, G., KRISTENSEN, D., AND RAHBK, A. Modeling corporate defaults: Poisson autoregressions with exogenous covariates (parx). *Journal of Empirical Finance* 38 (2016), 640–663. Recent developments in financial econometrics and empirical finance.
- [2] BHASKARAN, K., GASPARRINI, A., HAJAT, S., SMEETH, L., AND ARMSTRONG, B. Time series regression studies in environmental epidemiology. *International journal of epidemiology* 42, 4 (2013), 1187–1195.
- [3] BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31 (1986), 307–327. [MR0853051](#)
- [4] BOLLERSLEV, T. Glossary to arch (garch). Tech. rep., CREATES Research Paper, September 2008. [MR3237182](#)
- [5] BOUGEROL, P., AND PICARD, N. Stationarity of garch processes and of some nonnegative time series. *J. Econometrics* 52, 1992 (1992), 115–127. [MR1165646](#)
- [6] CHRISTOU, V., AND FOKIANOS, K. Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis* 35, 1 (2014), 55–78. [MR3148248](#)
- [7] CHRISTOU, V., AND FOKIANOS, K. On count time series prediction. *Journal of Statistical Computation and Simulation* 85, 2 (2015), 357–373. [MR3270681](#)
- [8] COX, D. Statistical analysis of time-series: some recent developments. *Scand. J. Statist.* 8, 2 (1981), 93–115. [MR0623586](#)
- [9] CUI, Y., AND ZHU, F. A new bivariate integer-valued garch model allowing for negative cross-correlation. *TEST* 27, 2 (Jun 2018), 428–452. [MR3799147](#)
- [10] DAVIS, R., DUNSMUIR, W., AND STRETT, S. Observation-driven models for Poisson counts. *Biometrika* 90, 4 (DEC 2003), 777–790. [MR2024757](#)
- [11] DAVIS, R., AND LIU, H. Theory and inference for a class of observation-driven models with application to time series of counts. *Preprint, arXiv:1204.3915* (2012). [MR3586234](#)
- [12] DAVIS, R. A., DUNSMUIR, W., AND WANG, Y. Modeling time series of count data. *Statistics textbooks and monographs* 158 (1999), 63–114. [MR1724696](#)
- [13] DOUC, R., DOUKHAN, P., AND MOULINES, E. Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stoch. Proc. App.* 123, 7 (2013), 2620–2647. [MR3054539](#)
- [14] DOUC, R., MOULINES, E., PRIOURET, P., AND SOULIER, P. *Markov chains*. Springer, 2018. [MR3889011](#)
- [15] DOUC, R., ROUEFF, F., AND SIM, T. Handy sufficient conditions for the convergence of the maximum likelihood estimator in observation-driven models. *Lithuanian Mathematical Journal* 55, 3 (2015), 367–392. [MR3379032](#)
- [16] DOUC, R., ROUEFF, F., AND SIM, T. The maximizing set of the asymptotic normalized log-likelihood for partially observed markov chains. *Annals of Applied Probability* 26, 4 (2016), 2357–2383. [MR3543899](#)

- [17] DOUC, R., ROUEFF, F., AND SIM, T. Necessary and sufficient conditions for the identifiability of observation-driven models. *Journal of Time Series Analysis* (2020). Preprint available at [\[HAL\]](#).
- [18] DOUC, R., ROUEFF, F., AND SIM, T. Supplement to “necessary and sufficient conditions for the identifiability of observation-driven models”. *Journal of Time Series Analysis* (2020).
- [19] DOUKHAN, P., FOKIANOS, K., AND TJØSTHEIM, D. On weak dependence conditions for Poisson autoregressions. *Statist. Probab. Lett.* 82, 5 (2012), 942–948. [MR2910041](#)
- [20] FERLAND, R., LATOUR, A., AND ORAICHI, D. Integer-valued GARCH process. *J. Time Ser. Anal.* 27, 6 (2006), 923–942. [MR2328548](#)
- [21] FOKIANOS, K., NEUMANN, M. H., ET AL. A goodness-of-fit test for poisson count processes. *Electronic Journal of Statistics* 7 (2013), 793–819. [MR3040560](#)
- [22] FOKIANOS, K., RAHBEK, A., AND TJØSTHEIM, D. Poisson autoregression. *J. Am. Statist. Assoc.* 104, 488 (2009), 1430–1439. With electronic supplementary materials available online. [MR2596998](#)
- [23] FOKIANOS, K., AND TJØSTHEIM, D. Log-linear poisson autoregression. *J. of Multivariate Analysis* 102, 3 (2011), 563–578. [MR2755016](#)
- [24] FOKIANOS, K., AND TJØSTHEIM, D. Nonlinear poisson autoregression. *Annals of the Institute of Statistical Mathematics* 64, 6 (2012), 1205–1225. [MR2981620](#)
- [25] FRANCO, C., AND ZAKOIAN, J.-M. Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli* 10, 4 (2004), 605–637. [MR2076065](#)
- [26] FRANCO, C., AND ZAKOIAN, J.-M. A tour in the asymptotic theory of garch estimation. *Handbook of Financial Time Series* (2009), 85–111.
- [27] FRANCO, C., AND ZAKOIAN, J.-M. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2011. [MR3186556](#)
- [28] FRANKE, J. *Weak dependence of functional INGARCH processes*. Technische Universität Kaiserslautern, Fachbereich Mathematik, 2010.
- [29] HEINEN, A. Modelling time series count data: an autoregressive conditional poisson model. *Available at SSRN 1117187* (2003).
- [30] HENDERSON, S. G., MATTESON, D., AND WOODARD, D. Stationarity of generalized autoregressive moving average models. *Electronic Journal of Statistics* 5 (2011), 800–828. [MR2824817](#)
- [31] IVES, A., DENNIS, B., COTTINGHAM, K., AND CARPENTER, S. Estimating community stability and ecological interactions from time-series data. *Ecological monographs* 73, 2 (2003), 301–330.
- [32] LEROUX, B. G. Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.* 40 (1992), 127–143. [MR1145463](#)
- [33] LIESENFELD, R., AND RICHARD, J.-F. Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of empirical finance* 10, 4 (2003), 505–531.

- [34] LINDNER, A. M. Stationarity, mixing, distributional properties and moments of garch  $(p, q)$ -processes. In *Handbook of financial time series*. Springer, 2009, pp. 43–69.
- [35] MEYN, S. P., AND TWEEDIE, R. L. *Markov Chains and Stochastic Stability*. Cambridge University Press, London, 2009. [MR2509253](#)
- [36] NEUMANN, M. H. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* 17, 4 (NOV 2011), 1268–1284. [MR2854772](#)
- [37] PFANZAGL, J. On the measurability and consistency of minimum contrast estimates. *Metrika* 14 (1969), 249–272. [MR0250949](#)
- [38] PINDYCK, R. S., AND RUBINFELD, D. L. *Econometric models and economic forecasts*, vol. 4. Irwin/McGraw-Hill Boston, 1998.
- [39] RYDBERG, T. H., AND SHEPHARD, N. Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics* 1, 1 (2003), 2–25.
- [40] SIM, T. *Maximum likelihood estimation in partially observed Markov models with applications to time series of counts*. Theses, Télécom ParisTech, Mar. 2016.
- [41] STRAUMANN, D., AND MIKOSCH, T. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *ArXiv Mathematics e-prints* (Feb. 2007). [MR2291507](#)
- [42] STRETT, S. *Some observation driven models for time series of counts*. PhD thesis, Colorado State University, Department of Statistics, 2000. [MR2701787](#)
- [43] SURVEY, U. G. Earthquake hazards program.
- [44] TJØSTHEIM, D. *Count Time Series with Observation-Driven Autoregressive Parameter Dynamics*. Handbook of Discrete-Valued Time Series. London: Chapman & Hall, 2015.
- [45] TWEEDIE, R. L. Invariant measures for Markov chains with no irreducibility assumptions. *Journal of Applied Probability* 25 (1988), 275–285. [MR0974587](#)
- [46] WANG, C., LIU, H., YAO, J.-F., DAVIS, R. A., AND LI, W. K. Self-excited threshold poisson autoregression. *Journal of the American Statistical Association* 109, 506 (2014), 777–787. [MR3223749](#)
- [47] ZEGER, S. L. A regression model for time series of counts. *Biometrika* 75, 4 (1988), 621–629. [MR0995107](#)
- [48] ZHU, F. A negative binomial integer-valued GARCH model. *J. Time Series Anal.* 32, 1 (2011), 54–67. [MR2790672](#)