



Necessary and sufficient conditions for the identifiability of observation-driven models

Randal Douc, François Roueff, Tepmony Sim

► To cite this version:

Randal Douc, François Roueff, Tepmony Sim. Necessary and sufficient conditions for the identifiability of observation-driven models. *Journal of Time Series Analysis*, 2021, 42 (2), pp.140-160. <10.1111/jtsa.12559>. <hal-02088860v3>

HAL Id: hal-02088860

<https://hal.science/hal-02088860v3>

Submitted on 8 Sep 2020

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

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



HAL Authorization

NECESSARY AND SUFFICIENT CONDITIONS FOR THE IDENTIFIABILITY OF OBSERVATION-DRIVEN MODELS

RANDAL DOUC, FRANÇOIS ROUEFF, AND TEPMONY SIM

ABSTRACT. In this contribution we are interested in proving that a given observation-driven model is identifiable. In the case of a $\text{GARCH}(p, q)$ model, a simple sufficient condition has been established in [2] for showing the consistency of the quasi-maximum likelihood estimator. It turns out that this condition applies for a much larger class of observation-driven models, that we call the class of linearly observation-driven models. This class includes standard integer valued observation-driven time series such as the Poisson autoregression model and its numerous extensions. Our results also apply to vector-valued time series such as the bivariate integer valued GARCH model, to non-linear models such as the threshold Poisson autoregression or to observation-driven models with exogenous covariates such as the PARX model.

1. INTRODUCTION

Observation-driven models (ODM) were introduced in [8] and have received considerable attention since. They are commonly used for modeling various non-linear times series in applications ranging from economics (see [23]), environmental study (see [3]), epidemiology and public health study (see [29, 10, 12]), finance (see [20, 24, 13, 16]) and population dynamics (see [19]). Additional covariates have been added to some of these models leading to GARCHX type models, see [1] for recent examples in the context of count data, and the references therein. We include such a case in our setting leading to the general observation-driven models with exogenous variables (ODMX).

As often for non-linear time series the question of identifiability of the observation-driven models is a delicate one and is often appearing as an assumption used for proving the consistency (say) of the maximum likelihood estimator. A noticeable exception is the $\text{GARCH}(p, q)$ model, for which an explicit sufficient condition appears in [2], see their condition (2.27). We will in fact prove that this condition is not only sufficient but also necessary for the identifiability, and that this result extends to a much larger class of observation-driven models than the $\text{GARCH}(p, q)$ model. See [Theorem 17](#) below and the comments following this result.

We provide general conditions to ensure that an ODM or an ODMX defined through a collection of parameterized iterative schemes uniquely describes the law of the observations. In other words our conditions ensure that two different iterative schemes within the same model cannot produce the same law for the observations. Then a given parameter is identifiable if two different values of the parameter are not compatible with the same iterative scheme. Let us stress, however, that we do not consider the misspecified case here, that is, we always assume that the observations indeed follow the (unique) stationary distribution corresponding to (at least) one given parameter of the model. Our setting is nevertheless of interest for

Date: September 8, 2020.

2000 Mathematics Subject Classification. Primary: 60J05, 62F12; Secondary: 62M05, 62M10.

Key words and phrases. identifiability, observation-driven models, time series of counts.

the misspecified setting since a non-identifiable parameter (in the well specified case) cannot be identified in the misspecified case. Hence the necessity of our conditions remains true for the misspecified setting.

A special class of ODMs, that we call *linearly observation driven models* (LODMs) below, arises when the hidden variable is obtained linearly from hidden or observed variables of the past, and when all these variables are univariate, as for the GARCH(p, q) model. This latter model was extensively studied, see for example [5, 14, 15, 21, 16] and the references therein. Many other examples, linear or non-linear, univariate or multivariate, have been derived from this class, see [4] for a long list of them, although this list have been lengthened quite significantly since, in particular because of the recent adding of various integer valued ODMs to deal with count time series (see [7, 25] and the references therein). Our goal is to derive necessary and sufficient conditions potentially applying to a wide variety of ergodic observation driven models. To illustrate the generality of our results, we apply them to a list of various examples which includes, in addition to the standard GARCH model, the nonlinear GARCH model of [17], the INGARCH model of [12], the Log-linear Poisson GARCH of [13], the MPINGARCH model of [25], the PARX model of [1], the Bi-variate integer GARCH model of [9] and the self-excited threshold Poisson Autoregression of [28]. We are able to derive necessary and sufficient conditions for the identifiability of all the considered models.

The rest of the paper is organized as follows. Section 2 contains additional notation and definitions that will be used throughout the paper. Section 3 contains a list of examples already considered in the literature. Our main results can be found in Section 4, some proofs of which are postponed to Appendix A. In Section 5, we show how our results apply to the examples of Section 3 or can be extended to larger classes of models.

2. PRELIMINARIES

2.1. Formal definitions of observation driven models. Let us now formally introduce the class of observation-driven models and important sub-classes. 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 $(x_{0:(-1)}, y) = y$. An observation-driven time series model can formally be defined as follows.

Definition 1 (ODM, ODMX). Let $(\mathbf{X}, \mathcal{X})$, $(\mathbf{Y}, \mathcal{Y})$ and $(\mathbf{U}, \mathcal{U})$ be measurable spaces, respectively called the *latent space*, the *observation space* and the *admissible observation space*. Let (Θ, Δ) be a compact metric space, called the *parameter space*. Let Υ be a measurable function from $(\mathbf{Y}, \mathcal{Y})$ to $(\mathbf{U}, \mathcal{U})$. 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 $(\mathbf{X}^p \times \mathbf{U}^q, \mathcal{X}^{\otimes p} \otimes \mathcal{U}^{\otimes q})$ to $(\mathbf{X}, \mathcal{X})$, called the *reduced link functions* and let $\{G^\theta : \theta \in \Theta\}$ be a family of probability kernels on $\mathbf{X} \times \mathcal{Y}$, called the *observation kernels*. A time series $\{Y_k : k \geq -q + 1\}$ valued in \mathbf{Y} is said to be distributed according to an *observation-driven model of order (p, q)* (hereafter, ODM(p, q)) with reduced link function $\tilde{\psi}^\theta$, admissible mapping Υ and observation kernel G^θ if there exists a process $\{X_k : k \geq -p + 1\}$ on $(\mathbf{X}, \mathcal{X})$ such that for all $k \in \mathbb{Z}_{\geq 0}$,

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

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

In the presence of exogenous variables defined as an r -Markov chain valued in the space $(\mathbf{V}, \mathcal{V})$ with kernel H , the admissible mapping Υ is defined from $\mathbf{Y} \times \mathbf{V}$ to

\mathbf{U} and the iterative equation (2.1) is replaced by, for all $k \in \mathbb{Z}_{\geq 0}$,

$$(2.2) \quad \begin{aligned} V_k &| \mathcal{F}_k \sim H(V_{(k-r):(k-1)}; \cdot), \\ Y_k &| \mathcal{F}_k \sim G^\theta(X_k; \cdot), \\ X_{k+1} &= \tilde{\psi}_{U_{(k-q+1):k}}^\theta(X_{(k-p+1):k}), \end{aligned}$$

where, in this case, $\mathcal{F}_k = \sigma(X_{(-p+1):k}, Y_{(-q+1):(k-1)}, V_{((-r) \wedge (-q+1)):(k-1)})$ and $U_j = \Upsilon(Y_j, V_j)$ for all $j > -q$. We then say that the time series $\{Y_k : k \geq -q+1\}$ valued in \mathbf{Y} is distributed according to an *observation-driven model of order (p, q) with r -order Markov exogenous variables $\{V_k : k \geq ((-r) \wedge (-q+1))\}$* (hereafter, $\text{ODMX}(p, q, r)$) with reduced link function $\tilde{\psi}^\theta$, admissible mapping Υ , observation kernel G^θ , and exogenous Markov kernel H .

The variables Y_k are called the *observed variables*, the variables X_k the *hidden variables* and the variables U_k the *admissible variables*. In addition, we define the *augmented variables*

$$(2.3) \quad Z_k = (X_{(k-p+1):k}, U_{(k-q+1):(k-1)}) \in \mathbf{Z},$$

which take values in the *augmented space*

$$(2.4) \quad \mathbf{Z} = \mathbf{X}^p \times \mathbf{U}^{q-1} \quad \text{endowed with the } \sigma\text{-field } \mathcal{Z} = \mathcal{X}^{\otimes p} \otimes \mathcal{U}^{\otimes (q-1)}.$$

Remark 1. Let us briefly comment on the unusual notion of *admissible mapping* which allows us to define the admissible variables $U_j = \Upsilon(Y_j)$ of an $\text{ODM}(p, q)$ in [Definition 1](#):

- (1) For all $k \geq 0$, the conditional distribution of (Y_k, X_{k+1}) given \mathcal{F}_k only depends on Z_k defined by (2.3).
- (2) The time series $\{U_k : k > -q\}$ is also an $\text{ODM}(p, q)$ with admissible mapping being the identity, link function $\tilde{\psi}^\theta$ and observation kernel $\tilde{G}^\theta(x, \cdot) = G^\theta(x, \Upsilon^{-1}(\cdot))$ on the observation space $(\mathbf{U}, \mathcal{U})$.
- (3) On the other hand, we can also set the admissible mapping to be the identity for the $\text{ODM} \{Y_k : k > -q\}$, in which case $\mathbf{Y} \subseteq \mathbf{U}$ and the reduced link function should be replaced by the *link function* defined all $(x, y_{1:q}) \in \mathbf{X}^p \times \mathbf{Y}^q$ by

$$(2.5) \quad \psi_{y_{1:q}}^\theta(x) = \tilde{\psi}_{u_{1:q}}^\theta(x) \quad \text{with} \quad u_k = \Upsilon(y_k) \quad \text{for} \quad 1 \leq k \leq q.$$

In fact the advantage of using an admissible mapping is precisely to obtain a reduced link function $\tilde{\psi}$, more convenient than the (non-reduced) link function ψ . We will focus in the particular case where $\tilde{\psi}$ is linear, into which we can cast not all but many observation driven models, see [Section 3](#) hereafter.

- (4) An $\text{ODMX}(p, q, r)$ can be cast into an $\text{ODM}(p, q)$ by defining $\tilde{Y}_k = (Y_k, V_{(k-r+1):k})$ and $\tilde{X}_k = (X_k, V_{(k-r):(k-1)})$ and observing that the obtained times series $\{\tilde{Y}_k : k \geq -q+1\}$ is an $\text{ODM}(p, q)$ with hidden variables $\{\tilde{X}_k : k \geq -p+1\}$. However for treating identifiability as is the purpose here, it is more convenient to keep distinguishing between the ODM and the ODMX setting.
- (5) In the following the variables U_k and Z_k will be used extensively as they simplify a lot the presentation and the reasoning. It is important to note that the definitions of U_k , Z_k and \mathcal{F}_k are not the same in the ODM and the ODMX settings as they involve V_k in the later case. In particular the conditional distribution of U_k given \mathcal{F}_k takes two very different forms in the ODM and ODMX cases. They can be respectively expressed by $\tilde{G}^\theta(X_k; \cdot)$ and $\tilde{G}^\theta((X_k, V_{(k-r):(k-1)}); \cdot)$ where \tilde{G}^θ is a probability kernel on $\mathbf{X} \times \mathbf{U}$ and

$(\mathbf{X} \times \mathbf{V}^r) \times \mathcal{U}$, resp. For conciseness we use the same notation \tilde{G}^θ for the two cases. They are respectively defined by setting, for all $x \in \mathbf{X}$, $A \in \mathcal{U}$ and $v \in \mathbf{V}^r$,

$$(2.6) \quad \tilde{G}^\theta(x, A) = G^\theta(x, \Upsilon^{-1}(A)) ,$$

$$(2.7) \quad \tilde{G}^\theta((x, v), A) = \int_{\mathbf{Y} \times \mathbf{V}} \mathbb{1}_A(\Upsilon(y, w)) G^\theta(x; dy) H(v; dw) .$$

When the reduced link function is linear we specify [Definition 1](#) into the following.

Definition 2 ((V)LODM(X)). We say that an ODM(p, q) (resp. ODMX(p, q, r)) is a *vector linearly observation-driven model of order* (p, q, p', q') , shortened as VLODM(p, q, p', q'), (resp. VLODMX(p, q, r, p', q')) if for some $p', q' \in \mathbb{Z}_{>0}$, \mathbf{X} and \mathbf{U} are closed subsets of $\mathbb{R}^{p'}$ and $\mathbb{R}^{q'}$, respectively, and, for all $x = x_{0:(p-1)} \in \mathbf{X}^p$, $u = u_{0:(q-1)} \in \mathbf{U}^q$, and $\theta \in \Theta$,

$$(2.8) \quad \tilde{\psi}_u^\theta(x) = \omega(\theta) + \sum_{i=1}^p A_i(\theta) x_{p-i} + \sum_{i=1}^q B_i(\theta) u_{q-i} ,$$

for some mappings ω , $A_{1:p}$ and $B_{1:q}$ defined on Θ 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 VLODM(p, q, p', q') (resp. VLODMX(p, q, r, p', q')) is simply called a *linearly observation-driven model of order* (p, q) , shortened as LODM(p, q) (resp. LODMX(p, q, r)).

2.2. Iterations of the link function. We now introduce iterated versions of the reduced link function $\tilde{\psi}^\theta$. Let \mathbf{Z} be defined by (2.4). We define for any $k \in \mathbb{Z}_{>0}$ and $u_{0:(k-1)} \in \mathbf{U}^k$, the mapping $\tilde{\psi}^\theta \langle u_{0:(k-1)} \rangle : \mathbf{Z} \rightarrow \mathbf{X}$ through a set of recursive equations of order (p, q) . Namely, for all $n \in \mathbb{Z}_{>0}$, $u_{0:(k-1)} \in \mathbf{U}^k$ and $z = z_{1:(p+q-1)} \in \mathbf{Z}$, we define

$$(2.9) \quad \tilde{\psi}^\theta \langle u_{0:(k-1)} \rangle(z) := x_k ,$$

where the sequence $x_{(-p+1):k}$ is defined by

$$(2.10) \quad \begin{cases} u_j = z_{p+q+j} , & -q < j \leq -1 , \\ x_j = z_{p+j} , & -p < j \leq 0 , \\ x_j = \tilde{\psi}_{u_{(j-q):(j-1)}}^\theta(x_{(j-p):(j-1)}) , & 1 \leq j \leq k . \end{cases}$$

In this set of equations the last line is applied recursively so that in fact, for all $j \geq 1$, x_j only depends on z and $u_{0:(j-1)}$.

The equations in (2.10) define a system with input sequence $u_{(-q+1):(k-1)}$, initial condition $x_{(-p+1):0}$ and output sequence $x_{1:k}$. Because the recursion given by the last line of (2.10) involves $p+1$ successive entries of the output and q successive entries of the input, it is useful to define blocks, valued in $\mathbf{Z} = \mathbf{X}^p \times \mathbf{U}^{q-1}$ and consider the same recursion applying to such blocks, hence computing z_j from z_{j-1} and u_{j-1} . Formally, for all $u \in \mathbf{U}$, we define $\tilde{\Psi}_u^\theta : \mathbf{Z} \rightarrow \mathbf{Z}$ by

$$(2.11) \quad \tilde{\Psi}_u^\theta : (x_{1:p}, u_{1:(q-1)}) \mapsto \begin{cases} (x_{2:p}, \tilde{\psi}_{u_{1:(q-1)}, u}^\theta(x_{1:p}), u_{2:(q-1)}, u) & \text{if } q > 1 \\ (x_{2:p}, \tilde{\psi}_u^\theta(x_{1:p})) & \text{if } q = 1 , \end{cases}$$

Remark 2. Note in particular that with this notation at hand, and using the admissible variables $U_k = \Upsilon(Y_k)$ for the ODM case or $U_k = \Upsilon(Y_k, V_k)$ for the ODMX case, and Z_k defined by (2.3), the second line of (2.1) and the third line of (2.2) are equivalent to

$$(2.12) \quad Z_{k+1} = \tilde{\Psi}_{U_k}^\theta(Z_k) .$$

We further denote the successive composition of $\tilde{\Psi}_{u_0}^\theta$, $\tilde{\Psi}_{u_1}^\theta$, ..., and $\tilde{\Psi}_{u_{k-1}}^\theta$ by

$$(2.13) \quad \tilde{\Psi}^\theta \langle u_{0:(k-1)} \rangle = \tilde{\Psi}_{u_{k-1}}^\theta \circ \tilde{\Psi}_{u_{k-2}}^\theta \circ \dots \circ \tilde{\Psi}_{u_0}^\theta.$$

This recursion is the same as the one for defining $\tilde{\psi}^\theta \langle u \rangle$, except that it is valued in \mathbb{Z} , whereas $\tilde{\psi}^\theta \langle u \rangle$ is valued in \mathbb{X} . More precisely, denoting, throughout the paper, for all $j \in \{1, \dots, p+q-1\}$, by $\Pi_j(z)$ the j -th entry of $z \in \mathbb{Z}$, we have the following relations between $\tilde{\psi}^\theta \langle u \rangle$ and $\tilde{\Psi}^\theta \langle u \rangle$, for all $k \in \mathbb{Z}_{\geq 0}$ and $u \in \mathbb{U}^k$,

$$(2.14) \quad \tilde{\psi}^\theta \langle u \rangle = \Pi_p \circ \tilde{\Psi}^\theta \langle u \rangle,$$

$$(2.15) \quad \tilde{\Psi}^\theta \langle u_{0:(k-1)} \rangle(z) = \left(\left(\tilde{\psi}^\theta \langle u_{0:j} \rangle(z) \right)_{k-p \leq j < k}, u_{(k-q+1):(k-1)} \right),$$

where, in the second line, we set $u_j = \Pi_{p+q+j}(z)$ for $-q < j \leq -1$ and use the convention $\tilde{\psi}^\theta \langle u_{0:j} \rangle(z) = \Pi_{p-j}(z)$ for $-p < j \leq 0$.

2.3. Ergodic assumption and some interesting class of parameters. In this contribution, we only consider the case where all processes in the model are ergodic. Namely, we use the following assumption.

(A-1) For all $\theta \in \Theta$, there exists a unique stationary solution $\{(X_k, Y_k) : k \in \mathbb{Z}\}$ satisfying (2.1).

In the case of exogenous covariates this assumption is replaced by the following.

(A'-1) For all $\theta \in \Theta$, there exists a unique stationary solution $\{(X_k, Y_k, V_k) : k \in \mathbb{Z}\}$ satisfying (2.2).

This ergodic property is the cornerstone for making statistical inference theory work and we provide simple general conditions in [11] for $p = q = 1$ and in [26, Chapter 5] and [27] for the case of general order (p, q) .

We now introduce the notation that will allow us to refer to the stationary distribution of the model throughout the paper.

Definition 3 (Stationary distributions \mathbb{P}^θ and $\tilde{\mathbb{P}}^\theta$). We define the distributions \mathbb{P}^θ and $\tilde{\mathbb{P}}^\theta$ as follows.

- a) Under (A-1), \mathbb{P}^θ denotes the distribution on $((\mathbb{X} \times \mathbb{Y})^\mathbb{Z}, (\mathcal{X} \times \mathcal{Y})^{\otimes \mathbb{Z}})$ of the stationary solution of (2.1); Under (A'-1), \mathbb{P}^θ denotes the distribution on $((\mathbb{X} \times \mathbb{Y} \times \mathbb{V})^\mathbb{Z}, (\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{V})^{\otimes \mathbb{Z}})$ of the stationary solution of (2.2).
- b) Under (A-1), $\tilde{\mathbb{P}}^\theta$ denotes the projection of \mathbb{P}^θ on the component $\mathbb{Y}^\mathbb{Z}$; Under (A'-1), $\tilde{\mathbb{P}}^\theta$ denotes the projection of \mathbb{P}^θ on the component $(\mathbb{Y} \times \mathbb{V})^\mathbb{Z}$.

We also use the symbols \mathbb{E}^θ and $\tilde{\mathbb{E}}^\theta$ to denote the expectations corresponding to \mathbb{P}^θ and $\tilde{\mathbb{P}}^\theta$, respectively.

To study the identifiability of ergodic ODM's, we introduce equivalent classes that define a partition of the parameter set in subsets of parameters which share the same distribution of observations. Formally, it reads as follows.

Definition 4 (Equivalent classes for $\tilde{\mathbb{P}}^\theta$). Suppose that (A-1) or (A'-1) holds and define $\tilde{\mathbb{P}}^\theta$ as in Definition 3. For all $\theta, \theta' \in \Theta$, we write $\theta \sim \theta'$ if and only if $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta'}$. This defines an equivalence relation on the parameter set Θ and, for any $\theta \in \Theta$, the equivalence class of θ is denoted by $[\theta] := \{\theta' \in \Theta : \theta' \sim \theta\}$.

Remark 3. In the context of exogenous variables, that is, under (A'-1), since the distribution of $\{V_k : k \in \mathbb{Z}\}$ under $\tilde{\mathbb{P}}^\theta$ does not depend on θ , $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta'}$ is equivalent to say that the conditional distribution of $\{Y_k : k \in \mathbb{Z}\}$ given $\{V_k : k \in \mathbb{Z}\}$ is the same under $\tilde{\mathbb{P}}^\theta$ and under $\tilde{\mathbb{P}}^{\theta'}$.

Determining the equivalent classes $[\theta]$ for all $\theta \in \Theta$ amounts to solve the identifiability of a parameter under the assumption of a *well specified* model. Namely, assuming that the distribution of the observations is given by $\tilde{\mathbb{P}}^{\theta_\star}$ for some (unknown) parameter $\theta_\star \in \Theta$, a parameter $\xi(\theta_\star)$ is identifiable if and only if the given mapping ξ is constant over the equivalent class $[\theta_\star]$. Without identifiability, the consistency of any estimator of $\xi(\theta_\star)$ is not possible. A special case is when $[\theta_\star]$ reduces to the singleton $\{\theta_\star\}$, so that every parameter $\xi(\theta_\star)$ is identifiable, in which case the model is said to be identifiable. Obviously, if θ and θ_\star share the same iterative equation (2.1) (or (2.2 with exogenous covariates), that is, if $G^\theta = G^{\theta_\star}$ and $\tilde{\psi}_u^\theta(x) = \tilde{\psi}_u^{\theta_\star}(x)$ for all $(u, x) \in U^q \times \mathbf{X}^p$, by uniqueness of the stationary distribution, they must share the same one and in particular we get $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta_\star}$. Thus, using the more convenient notation $\tilde{\Psi}$ introduced in (2.11), we have

$$(2.16) \quad \left\{ \theta \in \Theta : G^\theta = G^{\theta_\star} \text{ and } \tilde{\Psi}_u^\theta(z) = \tilde{\Psi}_u^{\theta_\star}(z) \text{ for all } (z, u) \in Z \times U \right\} \subseteq [\theta_\star].$$

We will provide general conditions ensuring that this inclusion becomes an equality, see Corollary 13 below. However it may happen in standard situations that this inclusion is strict, as will be seen in Remark 8(5). Nevertheless, in all the considered examples, it will be possible to recover an equality by replacing Z by a more appropriate subset in the left-hand side of (2.16).

As often for ODMs, our results rely on the assumption that, under \mathbb{P}^θ , the hidden variables are measurable with respect to the admissible variables from the past. This is not completely surprising since, using the notation introduced in Section 2.2, iterating the link function, we have that, for all $\theta \in \Theta$ and all $s < t$ in \mathbb{Z} ,

$$X_t = \tilde{\psi}^\theta \langle U_{s:(t-1)} \rangle (Z_s) \quad \mathbb{P}^\theta\text{-a.s.}$$

In particular, taking $t = 1$ and letting s decrease backward towards $-\infty$, we get that, X_1 is measurable with respect to $\cap_{t \in \mathbb{Z}} (\mathcal{F}_t^Z \vee \mathcal{F}_0^U)$, where (\mathcal{F}_t^Z) and (\mathcal{F}_t^U) respectively denote the natural filtrations of $\{Z_n : n \in \mathbb{Z}\}$ and $\{U_n : n \in \mathbb{Z}\}$. To our knowledge, all ODM of interest satisfy in fact the stronger property that X_1 is measurable with respect to \mathcal{F}_0^U , which is sometimes called the *invertibility condition*. This condition is now introduced with some notation for expressing X_1 as a measurable function of $U_{(-\infty):0}$.

(A-2) For all $\theta \in \Theta$, the measurable function $\tilde{\psi}^\theta \langle \cdot \rangle : U^{\mathbb{Z}_{\leq 0}} \rightarrow \mathbf{X}$ satisfies

$$(2.17) \quad X_1 = \tilde{\psi}^\theta \langle U_{(-\infty):0} \rangle \quad \mathbb{P}^\theta\text{-a.s.}$$

Since \mathbb{P}^θ is stationary, (2.17) also implies that, for all $t \in \mathbb{Z}$, $X_{t+1} = \tilde{\psi}^\theta \langle U_{(-\infty):t} \rangle$ \mathbb{P}^θ -a.s. For an ODM (resp. an ODMX), we have $U_k = \Upsilon(Y_k)$ (resp. $U_k = \Upsilon(Y_k, V_k)$). Thus Assumption (A-2) allows us to derive the X_t 's from the Y_t 's (resp. from the Y_t 's and V_t 's) and therefore to rewrite the relationship given through the link function in the second line of (2.1) (resp. in the third line of (2.2)) between these variables in terms of a recursive relationship involving only the Y_t 's (resp. the Y_t 's and the V_t 's). It turns out that Condition (2.17) in (A-2) can be verified using $\tilde{\mathbb{P}}^\theta$ only, that is, we do not need \mathbb{P}^θ but only its marginal onto the variable Y_k 's (resp. the variables Y_t 's and the V_t 's), as shown by the following result.

Lemma 5. *Consider an ODM(p, q) satisfying (A-1) with $p, q \in \mathbb{Z}_{>0}$ or an ODMX(p, q, r) satisfying (A'-1) with $p, q, r \in \mathbb{Z}_{>0}$. Let $\theta \in \Theta$ and consider a measurable function $\tilde{\psi}^\theta \langle \cdot \rangle : U^{\mathbb{Z}^-} \rightarrow \mathbf{X}$. Then (2.17) is satisfied if and only if the two*

following equations hold.

$$(2.18) \quad \tilde{\psi}^\theta \langle U_{(-\infty):0} \rangle = \tilde{\psi}_{U_{(-q+1):0}}^\theta \left(\left(\tilde{\psi}^\theta \langle U_{(-\infty):j} \rangle \right)_{-p \leq j \leq -1} \right) \quad \tilde{\mathbb{P}}^\theta\text{-a.s.}$$

$$(2.19) \quad \tilde{\mathbb{P}}^\theta [Y_1 \in \cdot | Y_{(-\infty):0}] = G^\theta \left(\tilde{\psi}^\theta \langle U_{(-\infty):0} \rangle, \cdot \right) \quad \tilde{\mathbb{P}}^\theta\text{-a.s.}$$

Proof. Suppose that (2.17) holds true. Since \mathbb{P}^θ is shift invariant, it can be extended to all time instants $k \in \mathbb{Z}$, namely,

$$X_k = \tilde{\psi}^\theta \langle U_{(-\infty):(k-1)} \rangle \quad \mathbb{P}^\theta\text{-a.s.}$$

But then (2.18) and (2.19) follows from the model equations (2.1) in the case of an ODM or (2.2) in the case of an ODMX.

Suppose now that (2.18) and (2.19) hold true. Since \mathbb{P}^θ is shift invariant, they are extended to all time instants $k \in \mathbb{Z}$ in the form

$$\tilde{\psi}^\theta \langle U_{(-\infty):k-1} \rangle = \tilde{\psi}_{U_{(k-q):(k-1)}}^\theta \left(\left(\tilde{\psi}^\theta \langle U_{(-\infty):j} \rangle \right)_{k-p-1 \leq j \leq k-2} \right) \quad \tilde{\mathbb{P}}^\theta\text{-a.s.}$$

$$\tilde{\mathbb{P}}^\theta [Y_k \in \cdot | Y_{(-\infty):(k-1)}] = G^\theta \left(\tilde{\psi}^\theta \langle U_{(-\infty):(k-1)} \rangle, \cdot \right) \quad \tilde{\mathbb{P}}^\theta\text{-a.s.}$$

Defining $X'_k = \tilde{\psi}^\theta \langle U_{(-\infty):(k-1)} \rangle$ for all $k \in \mathbb{Z}$, we see that $\{(X'_k, Y_k) : k \in \mathbb{Z}\}$ is a stationary sequence satisfying the model equations (2.1) in the ODM case and $\{(X'_k, Y_k, V_k) : k \in \mathbb{Z}\}$ is a stationary sequence satisfying the model equations (2.2) in the ODMX case. By uniqueness of \mathbb{P}^θ assumed in (A-1) and (A'-1), respectively, we get that (2.17) holds. \square

Now, given $\theta_\star \in \Theta$, we introduce the set $\langle \theta_\star \rangle$ of all parameters $\theta \in \Theta$ whose recursive relationship (2.18) apply to almost all trajectories of $\{U_n : n \in \mathbb{Z}\}$ under the distribution of θ_\star .

Definition 6 (Subset $\langle \theta_\star \rangle$). Suppose that we are given a measurable function $\tilde{\psi}^\theta \langle \cdot \rangle : U^{\mathbb{Z}_{\leq 0}} \rightarrow \mathbf{X}$. Then, for all $\theta_\star \in \Theta$, we denote by $\langle \theta_\star \rangle$ the set of all parameters $\theta \in \Theta$ satisfying the two following equations

$$(2.20) \quad \tilde{\psi}^\theta \langle U_{(-\infty):0} \rangle = \tilde{\psi}^{\theta_\star} \langle U_{(-\infty):0} \rangle \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

$$(2.21) \quad \tilde{\psi}^\theta \langle U_{(-\infty):0} \rangle = \tilde{\psi}_{U_{(-q+1):0}}^\theta \left(\left(\tilde{\psi}^\theta \langle U_{(-\infty):j} \rangle \right)_{-p \leq j \leq -1} \right) \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

It is important to note that $\langle \theta_\star \rangle$ of Definition 6 depends on the choice of the class of functions $\{\tilde{\psi}^\theta \langle \cdot \rangle : \theta \in \Theta\}$ and that Assumption (A-2) alone is not sufficient to define each $\tilde{\psi}^\theta \langle \cdot \rangle$ on the whole set $U^{\mathbb{Z}_{\leq 0}}$ of trajectories, since Relation (2.17) is only required to hold \mathbb{P}^θ -a.s.. We now provide some Lipschitz condition on the iterates of the link function $\tilde{\psi}^\theta$ and a moment condition on U_0 that allow us to build a natural class of functions $\{\tilde{\psi}^\theta \langle \cdot \rangle : \theta \in \Theta\}$ that satisfies (A-2). Whenever we need some metric on the space \mathbf{Z} , we assume the following.

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

Recall that, for any finite \mathbf{U} -valued sequence u , the mapping $\tilde{\psi}^\theta \langle u \rangle$ is defined by (2.9) following the recursion in (2.10). Define, for all $n \in \mathbb{Z}_{>0}$, the Lipschitz constant for $\tilde{\psi}^\theta \langle u \rangle$, uniform over $u \in U^n$,

$$(2.22) \quad \text{Lip}_n^\theta = \sup \left\{ \frac{\delta_{\mathbf{X}}(\tilde{\psi}^\theta \langle u \rangle(z), \tilde{\psi}^\theta \langle u \rangle(z'))}{\delta_{\mathbf{Z}}(z, z')} : (z, z', u) \in \mathbf{Z}^2 \times U^n \right\},$$

where we set, for all $v = v_{1:(p+q-1)} \in \mathbb{Z}$ and $v' = v'_{1:(p+q-1)} \in \mathbb{Z}$,

$$(2.23) \quad \delta_{\mathbb{Z}}(v, v') = \left(\max_{1 \leq k \leq p} \delta_{\mathbb{X}}(v_k, v'_k) \right) \vee \left(\max_{p < k \leq p+q} \delta_{\mathbb{U}}(v_k, v'_k) \right).$$

We use the following assumptions to define the class of functions $\{\tilde{\psi}^\theta \langle \cdot \rangle : \theta \in \Theta\}$.

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

(A-5) There exists $x_1^{(i)} \in \mathbb{X}$ and, if $q > 1$, $u_1^{(i)} \in \mathbb{U}$ such that the constant vectors $x^{(i)} = (x_1^{(i)}, \dots, x_1^{(i)}) \in \mathbb{X}^p$ and $u^{(i)} = (u_1^{(i)}, \dots, u_1^{(i)}) \in \mathbb{U}^{q-1}$ satisfy, for all $\theta_*, \theta \in \Theta$,

$$(2.24) \quad \mathbb{E}^{\theta_*} [\phi^\theta(U_0)] < \infty,$$

where we defined, for all $u \in \mathbb{Y}$,

$$\phi^\theta(u) = \ln^+ \left(\delta_{\mathbb{X}} \left(x_1^{(i)}, \tilde{\psi}_{(u^{(i)}, u)}^\theta(x^{(i)}) \right) \vee \delta_{\mathbb{U}}(u_1^{(i)}, u) \right)$$

with the convention $\delta_{\mathbb{U}}(u_1^{(i)}, u) = 0$ if $q = 1$.

(A-6) For all $\theta \in \Theta$ and $u \in \mathbb{U}^q$, the reduced link function $\tilde{\psi}_u^\theta$ is continuous on \mathbb{X}^p .

Obviously, under (A-4), for all $\theta \in \Theta$ and $u \in \mathbb{U}^{\mathbb{Z} \leq 0}$, the asymptotic behavior of $\tilde{\psi}^\theta \langle u_{(-n):0} \rangle(z)$ as $n \rightarrow \infty$ does not depend on $z \in \mathbb{Z}$. We can thus denote

$$(2.25) \quad \begin{cases} D^\theta &:= \left\{ u \in \mathbb{U}^{\mathbb{Z} \leq 0} : \tilde{\psi}^\theta \langle u_{(-n):0} \rangle(z) \text{ converges in } \mathbb{X} \text{ as } n \rightarrow \infty \right\}, \\ \tilde{\psi}^\theta \langle u \rangle &:= \lim_{n \rightarrow \infty} \tilde{\psi}^\theta \langle u_{(-n):0} \rangle(z) \text{ for all } u \in D^\theta, \end{cases}$$

and keep in mind that the initial point z has no influence on these two definitions. By (2.15), we can rewrite (2.25) as

$$(2.26) \quad \begin{cases} D^\theta &:= \left\{ u \in \mathbb{U}^{\mathbb{Z} \leq 0} : \tilde{\Psi}^\theta \langle u_{(-n):0} \rangle(z) \text{ converges in } \mathbb{Z} \text{ as } n \rightarrow \infty \right\}, \\ \tilde{\Psi}^\theta \langle u \rangle &:= \lim_{n \rightarrow \infty} \tilde{\Psi}^\theta \langle u_{(-n):0} \rangle(z) \text{ for all } u \in D^\theta, \end{cases}$$

where $\tilde{\Psi}^\theta \langle u \rangle$ and $\tilde{\psi}^\theta \langle u \rangle$ are related for all $u \in D^\theta$ through the formulas

$$(2.27) \quad \tilde{\psi}^\theta \langle u \rangle = \Pi_p \circ \tilde{\Psi}^\theta \langle u \rangle,$$

$$(2.28) \quad \tilde{\Psi}^\theta \langle u \rangle = \left(\left(\tilde{\psi}^\theta \langle u_{(-\infty):k} \rangle \right)_{-p < k \leq 0}, u_{(-q+2):0} \right).$$

Based on these definitions, we now introduce subsets of \mathbb{Z} of particular interest.

Definition 7 (Set E^θ). If Assumption (A-4) holds, we set, for any $\theta \in \Theta$,

$$(2.29) \quad E^\theta := \left\{ \tilde{\Psi}^\theta \langle u \rangle : u \in D^\theta \right\} \subset \mathbb{Z},$$

where $\tilde{\Psi}^\theta \langle \cdot \rangle$ and D^θ are defined by (2.26).

Remark 4. Suppose that, for all $\theta \in \Theta$, we have $U_{(-\infty):0} \in D^\theta$, $\tilde{\mathbb{P}}^\theta$ -a.s., and suppose that (A-2) holds for $\tilde{\psi}^\theta \langle \cdot \rangle$ as in (2.25). Then, by (2.3) and (2.28), we have $Z_1 \in E^\theta$, \mathbb{P}^θ -a.s. Since \mathbb{P}^θ is shift-invariant, we get that $\{Z_k : k \in \mathbb{Z}\}$ takes its values in E^θ , \mathbb{P}^θ -a.s. This is why the set E^θ will be of interest in the following.

The following result is proved in Section A.1.

Lemma 8. Consider an ODM(p, q) satisfying (A-1) with $p, q \in \mathbb{Z}_{>0}$ or an ODMX(p, q, r) satisfying (A'-1) with $p, q, r \in \mathbb{Z}_{>0}$. Suppose that (A-3), (A-4) and (A-5) hold. Then, for all $\theta, \theta_* \in \Theta$, we have $U_{(-\infty):0} \in D^\theta$, $\tilde{\mathbb{P}}^{\theta_*}$ -a.s., (A-2) holds and, setting E^{θ_*} as in Definition 7, we have

$$(2.30) \quad \left\{ \theta \in \Theta : G^\theta = G^{\theta_*} \text{ and } \tilde{\Psi}_u^\theta(z) = \tilde{\Psi}_u^{\theta_*}(z) \text{ for all } (z, u) \in E^{\theta_*} \times \mathbb{U} \right\} \subseteq [\theta_*].$$

If moreover (A-6) is assumed, then (2.21) holds for all $\theta \in \Theta$. Consequently, the set $\langle \theta_\star \rangle$ in Definition 6 can be expressed as

$$(2.31) \quad \langle \theta_\star \rangle = \left\{ \theta \in \Theta : \tilde{\psi}^\theta \langle U_{(-\infty):0} \rangle = \tilde{\psi}^{\theta_\star} \langle U_{(-\infty):0} \rangle \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \right\}.$$

Remark 5. The invertibility Assumption (A-2) is essential for deriving the identifiability class $[\theta_\star]$ using the set $\langle \theta_\star \rangle$. Lemma 8 can be used to prove it in all the examples that are considered hereafter. Indeed as will be checked in Section 5, all the considered examples satisfy the following facts:

- (1) The sets \mathbf{X} and \mathbf{U} are closed subsets of finite dimensional normed spaces and (A-3) follows.
- (2) Assumption (A-4) is weaker than what is needed for proving the ergodicity assumption (A-1). Consider for instance the classical GARCH(1,1) model defined by setting $\Upsilon(y) = y^2$, $\tilde{\psi}_u^\theta(x) = \omega + ax + bu$ and $G^\theta(x, \cdot) = \mathbb{P}(x\varepsilon \in \cdot)$ where ε is centered with variance 1. Then it is easily seen that (A-4) is equivalent to $a < 1$. On the other hand, the Lyapunov condition to get (A-1) reads $\mathbb{E} \log(b\varepsilon^2 + a) < 0$, which implies $a < 1$.
- (3) The moment condition (A-5) is implied by $\mathbb{E}^{\theta_\star} [\log^+(|Y_0|)] < \infty$, where $|\cdot|$ is some norm, and this condition holds as a byproduct of the proof of (A-1) (which often imply $\mathbb{E}^{\theta_\star} [|Y_0|^s] < \infty$ for some $s > 0$).
- (4) One can readily checks (A-6).

Note also that the set in the left-hand side of (2.30) contains the set in the left-hand side of (2.16). In all our examples, the assumptions of 13 below will be shown to hold, implying that the inclusion in (2.30) is in fact an equality. In some of these examples, however, the inclusion in (2.16) is strict, showing that the sets in the left-hand sides of (2.16) and (2.30) may happen to be different.

3. EXAMPLES

We give a non-exhaustive list of possible examples related to the previous definitions and for which our results apply, as will be shown in Section 5.

3.1. Standard LODMs. Many models can be considered as an LODM by choosing an appropriate admissible mapping Υ .

GARCH. The standard GARCH(p, q) model is a special case of LODM(p, q), in which case $\mathbf{X} = \mathbb{R}_{\geq 0}$, $\mathbf{Y} = \mathbb{R}$, $\Upsilon(y) = y^2$, and $G^\theta(x, \cdot)$ is a centered distribution with variance x , most commonly the normal distribution.

INGARCH. The standard Poisson integer-valued GARCH (INGARCH, see e.g. [12]) obviously is an LODM(p, q) with $\mathbf{X} = \mathbb{R}_{\geq 0}$, $\mathbf{Y} = \mathbb{Z}_{\geq 0}$ and $G^\theta(x, \cdot)$ is the Poisson distribution with mean x .

Extensions of INGARCH. Many extensions of the INGARCH model simply consist in extending the Poisson distribution to more general ones: the NBIN-GARCH model of [30], the COM-Poisson INGARCH model of [31], the zero-inflated Poisson GARCH of [32], or the mixed-Poisson integer GARCH (MPINGARCH) of [25], among others. Often for these extensions, an extra-parameter is used to define the distribution $G^\theta(x, \cdot)$, in which case this extra parameter can be taken either as known, in which case G^θ does not depend on θ , or as unknown, in which case G^θ only depends on a subparameter of θ . Some integer valued observation driven models require using a non-identity admissible mapping in order to be seen as an LODM. For instance, the log-linear Poisson Garch model of [13] is an LODM(p, q) by taking $\Upsilon(y) = \ln(1 + y)$, and $G^\theta(x, \cdot)$ as the Poisson distribution with mean e^x .

All the above examples are LODMs with a similar parametrization of the linear link function. In fact they only differ through the admissible mapping Υ or the observation kernel G^θ . We assemble them using the following definition.

Definition 9 (Standard LODM (with unknown observation kernel)). An LODM(p, q) of Definition 2 is said to be *standard* if $\theta = (\omega, a_{1:p}, b_{1:q}) \in \Theta \subset \mathbb{R}^{1+p+q}$ with $\omega(\theta) = \omega$, $A_k(\theta) = a_k$ for all $1 \leq k \leq p$ and $B_k(\theta) = b_k$ for all $1 \leq k \leq q$, and G^θ does not depend on θ , in which case we denote it by G . It is said to be *standard with unknown observation kernel* if the same holds with $\theta = (\vartheta, \varphi) \in \Theta \subset \mathbb{R}^{1+p+q} \times \Phi$ where $\vartheta = (\omega, a_{1:p}, b_{1:q})$ and Φ is some parameter set, and G^θ only depends on φ , in which case we denote it by G^φ .

In this definition the parameter φ is used in the case where the observation kernel depends on an unknown extra parameter, as considered in [25] for the class of MPINGARCH(p, q) models which include the NBIN GARCH model. A necessary and sufficient condition for standard LODMs with known or unknown observation kernel is provided in Theorem 17 below and applies to all the examples listed in this section.

3.2. A bivariate example. Let us extend Definition 9 to the vector case as follows.

Definition 10 (Standard VLODM (with unknown observation kernel)). A VLODM(p, q, p', q') is said to be *standard* if $\theta = (\omega, A_{1:p}, B_{1:q}) \in \Theta \subset \mathbb{R}^{p'} \times \left(\mathbb{R}^{p' \times p'}\right)^p \times \left(\mathbb{R}^{p' \times q'}\right)^q$ with $\omega(\theta) = \omega$, $A_k(\theta) = A_k$ for all $1 \leq k \leq p$ and $B_k(\theta) = B_k$ for all $1 \leq k \leq q$, and G^θ does not depend on θ , in which case we denote it by G . It is said to be *standard with unknown observation kernel* if the same holds with $\theta = (\vartheta, \varphi) \in \Theta \subset \mathbb{R}^{p'} \times \left(\mathbb{R}^{p' \times p'}\right)^p \times \left(\mathbb{R}^{p' \times q'}\right)^q \times \Phi$ where $\vartheta = (\omega, A_{1:p}, B_{1:q})$ and Φ is some parameter set, and G^θ only depends on φ , in which case we denote it by G^φ .

Then the bivariate integer valued GARCH model of [9] is a *standard* VLODM($1, 1, 2, 2$) with unknown observation kernel defined for all $\varphi \in \Phi = [-\bar{\varphi}, \bar{\varphi}]$ (where $\bar{\varphi} > 0$ is some constant), $(x_1, x_2) \in \mathbb{R}_{>0}$ and $(y_1, y_2) \in \mathbb{Z}_{\geq 0}$, by

$$(3.1) \quad G^\varphi((x_1, x_2), \{y_1, y_2\}) = \frac{x_1^{y_1} x_2^{y_2}}{y_1! y_2!} e^{-(x_1+x_2)} \left(1 + \varphi \prod_{i=1,2} (e^{-y_i} - e^{-c x_i}) \right).$$

where $c = 1 - 1/e$. Since we have $p = q = 1$ in this example, we simply denote $\theta = (\omega, A, B, \varphi) \in \mathbb{R}^2 \times \mathbb{R}^{2 \times 2} \times \mathbb{R}^{2 \times 2} \times [-\bar{\varphi}, \bar{\varphi}]$.

3.3. Non-linear GARCH. The non-linear GARCH model of [17] is an ODM(p, q) with

$$G^\theta(x, \cdot) = \mathbb{P}(x^{1/\delta} \eta \in \cdot)$$

$$\psi_{y_{0:(q-1)}}^\theta(x_{0:(p-1)}) = \omega + \sum_{i=1}^p a_i x_{p-i} + \sum_{i=1}^q (b_i(1)(y_{q-i}^+)^{\delta} + b_i(2)(y_{q-i}^-)^{\delta}),$$

where η is a real valued random variable. Two cases are considered in [17] :

- Case 1) If the exponent δ is known, we set $\theta = (\omega, a_{1:p}, \mathbf{b}_{1:q})$ with $\mathbf{b}_k = [b_k(1) \ b_k(2)]$ for $k = 1, \dots, q$, and $\Upsilon(y) = ((y^+)^{\delta}, (y^-)^{\delta})$, in which case we have a standard VLODM($p, q, 1, 2$) of Definition 10 with known observation kernel and with the parameters A_k denoted by a_k for $k = 1, \dots, p$ and the parameters B_k denoted by \mathbf{b}_k for $k = 1, \dots, q$.
- Case 2) If the exponent δ is unknown, we set $\theta = (\omega, a_{1:p}, \mathbf{b}_{1:q}, \delta)$ and $\Upsilon(y) = (y^+, y^-)$, in which case δ must be included in the definition of $\tilde{\psi}^\theta$.

To our best knowledge this kind of model have not be extended to the case of (signed) integer valued time series.

3.4. The SETPAR model. Other non-linear ODM's that cannot be cast into an LODM or a VLODM can be found in [6]. We consider here the *self-excited threshold Poisson autoregression* (SETPAR) model originally studied in [28], which is an ODM(1,1), integer valued ($Y = \mathbb{Z}_{\geq 0}$), with link function defined for all $\theta = (\omega_1, \omega_2, a_1, a_2, b_1, b_2, r) \in \Theta \subset \mathbb{R}_{\geq 0}^6 \times \mathbb{Z}_{\geq 0}$ by

$$(3.2) \quad \psi_y^\theta(x) = \begin{cases} \omega_1 + a_1 x + b_1 y & \text{if } y \leq r \\ \omega_2 + a_2 x + b_2 y & \text{if } y > r, \end{cases}$$

with $G^\theta(x, \cdot)$ being the usual Poisson distribution with mean x .

3.5. The PARX model. Our last example is the Poisson autoregression with exogenous covariates (PARX) model of [1]. The PARX model is similar to the standard INGARCH(p, q) model above but with additional exogenous variables entering into the link function for generating the hidden variables. The exogenous variables are assumed to satisfy some Markov dynamic of order 1 (see [1, Assumption 1]). Thus it is an ODMX($p, q, 1$). Eq. (1) in [1] corresponds to setting our $G^\theta(x, \cdot)$ as the Poisson distribution with mean x . Eq. (2) in [1] corresponds to setting for all $x = x_{0:(p-1)} \in \mathbb{R}^p$ and $u = (y_{0:(q-1)}, v) \in \mathbb{R}^q \times \mathbb{V}$,

$$(3.3) \quad \tilde{\psi}_u(x) = \omega + \sum_{k=1}^p a_k x_{p-k} + \sum_{k=1}^q b_k y_{q-k} + f(v, \gamma),$$

where $f(\cdot, \gamma) : \mathbb{V} \rightarrow \mathbb{R}_{\geq 0}$ is a known function and $\theta = (\omega, a_1, \dots, a_p, b_1, \dots, b_q, \gamma)$ is the unknown parameter of the model. Note that our $Y_k, X_k, V_k, a_{1:p}$ and $b_{1:q}$ correspond to their $y_k, \lambda_k, x_k, \beta_{1:q}$ and $\alpha_{1:p}$, respectively. Identifiability is considered in [1] by specifying γ as $\gamma = \gamma_{1:d} \in \mathbb{R}_{\geq 0}^d$ for some positive integer d (which corresponds to d_x in [1]) and $f(v, \gamma)$ as being of the form

$$(3.4) \quad f(v, \gamma) = \sum_{i=1}^d \gamma_i f_i(v),$$

for some known functions $f_1, \dots, f_d : \mathbb{V} \rightarrow \mathbb{R}_{\geq 0}$. It is in fact imposed in [1] that $v = v_{1:d} \in \mathbb{R}^d = \mathbb{V}$ and $f_i(v)$ actually is a function of v_i for each $i \in \{1, \dots, d\}$ but this constraint can be dropped for achieving wider generality without additional theoretical difficulties. The specific form of $f(v, \gamma)$ in (3.4) amounts in our setting to specify the previous ODMX($p, q, 1$) with reduced link function as in (3.3) to a VLODMX($p, q, 1, 1, d+1$) with $\Upsilon(y, v) = (y, f_1(v), \dots, f_d(v)) \in \mathbb{U} = \mathbb{R}^{1+d}$, $A_k(\theta) = a_k$ for $k = 1, \dots, p$, $B_1(\theta) = [b_1 \ \gamma_1 \ \dots \ \gamma_d]$ and $B_k(\theta) = [b_k \ 0 \ \dots \ 0]$ for $k = 2, \dots, q$. Then $\theta = (\omega, a_1, \dots, a_p, b_1, \dots, b_q, \gamma)$ with $\gamma = \gamma_{1:d} \in \mathbb{R}_{\geq 0}^d$ and it follows that Θ is a subset of $\mathbb{R}_{\geq 0}^{1+p+q+d}$.

4. MAIN RESULTS

4.1. General setting. To investigate the identifiability of the model, we first introduce an assumption which says how much can be identified from a single observation of the conditional distribution $G^\theta(x, \cdot)$.

(A-7) For all $\theta_\star \in \Theta$ there exists $[\theta_\star]_G \subset \Theta$ such that, for all $\theta \in \Theta$ and $x, x' \in \mathbb{X}$,

$$G^\theta(x; \cdot) = G^{\theta_\star}(x'; \cdot) \quad \text{if and only if} \quad \theta \in [\theta_\star]_G \quad \text{and} \quad x = x'.$$

It can be convenient to write the parameters as $\theta = (\vartheta, \varphi)$ so that G^θ only depends on φ , hence can be denoted by G^φ , and the link function ψ^θ only depends on ϑ , hence can be denoted by ψ^ϑ . In this case, the “if” in (A-7) holds by setting $[\theta_\star]_G = \{(\vartheta, \varphi) \in \Theta : \varphi = \varphi_\star\}$ for $\theta_\star = (\vartheta_\star, \varphi_\star)$, and the “only if” in (A-7) says

that $(\varphi, x) \mapsto G^\varphi(x, \cdot)$ is one-to-one. In many examples G^θ does not depend on θ at all, in which case $[\theta_\star]_G = \Theta$. See (SL'-3) below in Section 5.1 for such a case.

Our approach to establish identifiability is given by the following general result.

Proposition 11. *Consider an ODM(p, q) satisfying (A-1) with $p, q \in \mathbb{Z}_{>0}$ or an ODMX(p, q, r) satisfying (A'-1) with $p, q, r \in \mathbb{Z}_{>0}$. Let $\{\tilde{\psi}^\theta(\cdot) : \theta \in \Theta\}$ be a class of $U^{\mathbb{Z}_{\leq 0}} \rightarrow X$ -measurable functions satisfying (A-2). Suppose moreover that (A-7) holds. Then, for all $\theta_\star \in \Theta$, we have*

$$[\theta_\star] = [\theta_\star]_G \cap \langle \theta_\star \rangle,$$

where $\langle \theta_\star \rangle$, $[\theta_\star]_G$ and $[\theta_\star]$ are respectively defined in Definition 6, Assumption (A-7) and Definition 4.

The proof is postponed to Section A.2 for convenience. We now derive the main result of this section, which provides sufficient conditions in order to fully describe the set $\langle \theta_\star \rangle$. To this end, we introduce the following assumption, in which, by saying that a probability measure μ on (U, \mathcal{U}) is *non-degenerate* with respect to the class $\mathcal{C} \subset \mathcal{U}$, we mean that, for any $A \in \mathcal{C}$, $\mu(A) = 1$ can only be true if $A = U$.

(A-8) For all $\theta \in \Theta$ and $x \in X$, the measure $\tilde{G}^\theta(x; \cdot)$ defined by (2.6) on (U, \mathcal{U}) is non-degenerate with respect to the class \mathcal{C}^θ ,

where \mathcal{C}^θ denotes the class containing all sets $A \in \mathcal{U}$ for which there exist $\theta, \theta' \in \Theta$, $z \in Z$, $k, l \in \mathbb{Z}_{\geq 0}$, and $v, w \in U^k \times U^l$ such that

$$A = \left\{ u \in U : \tilde{\psi}^\theta \langle (v, u, w) \rangle(z) = \tilde{\psi}^{\theta'} \langle (v, u, w) \rangle(z) \right\}.$$

Remark 6. The non-degenerate assumption (A-8) is easy to check in the two following cases.

- (1) If for all $\theta \in \Theta$, $x \in X$ and $u \in U$, we have $\tilde{G}^\theta(x, \{u\}) > 0$, then for any set $A \in \mathcal{U}$ we have $\tilde{G}^\theta(x, A) = 1$ if and only if $A = U$. Thus (A-8) is immediately satisfied.
- (2) In the VLDM case, that is, with reduced link function given by (2.8), we immediately see that \mathcal{C}^θ only contains affine subsets (being the null space of an affine function). Hence we only need to require that $\tilde{G}^\theta(x; \cdot)$ does not have full measure on affine hyperplanes to ensure that it is non-degenerate with respect to the class \mathcal{C}^θ .

In the case of an ODMX, the definition of \tilde{G}^θ is different, see Remark 1(5), and Assumption (A-8) has to be adapted into the following.

(A'-8) For all $\theta \in \Theta$ and $w \in X \times V^r$, the measure $\tilde{G}^\theta(w; \cdot)$ defined by (2.7) on (U, \mathcal{U}) is non-degenerate with respect to the class \mathcal{C}^θ .

We have the following result.

Theorem 12. *Consider an ODM(p, q) satisfying (A-1) and (A-8) with $p, q \in \mathbb{Z}_{>0}$ or an ODMX(p, q, r) satisfying (A'-1) and (A'-8) with $p, q, r \in \mathbb{Z}_{>0}$. Assume that (A-3)–(A-6) hold. For all $\theta \in \Theta$, define D^θ and $\tilde{\psi}^\theta(\cdot)$ by (2.25). Then (A-2) holds and we have, for all $\theta_\star \in \Theta$,*

$$(4.1) \quad \langle \theta_\star \rangle = \left\{ \theta \in \Theta : \tilde{\Psi}_u^\theta(z) = \tilde{\Psi}_u^{\theta_\star}(z) \text{ for all } (z, u) \in E^{\theta_\star} \times U \right\},$$

where $\langle \theta_\star \rangle$ and E^{θ_\star} are as in Definition 6 and Definition 7, respectively.

Proof. The fact that (A-2) holds for the given choice of $\tilde{\psi}^\theta(\cdot)$ follows from Lemma 8.

Let us now take $\theta_\star \in \Theta$ and $\theta \in \langle \theta_\star \rangle$ and show that θ belongs to the right-hand side of (4.1). We prove this in the case of an ODMX satisfying (A'-1). (The case

of an ODM is readily obtained by removing the variables V_k 's in the reasoning). By (2.17), (2.20) and (2.21), and since $\mathbb{P}^{\theta_\star}$ is stationary, we have, for all $t \in \mathbb{Z}$,

$$X_{t+1} = \tilde{\psi}_{U_{(t-q+1):t}}^{\theta_\star}(X_{(t-p+1):t}) = \tilde{\psi}_{U_{(t-q+1):t}}^\theta(X_{(t-p+1):t}) \quad \mathbb{P}^{\theta_\star}\text{-a.s.}$$

Using (2.3) and the notation introduced in Section 2.2, we get that, for any $n \in \mathbb{Z}_{\geq 0}$,

$$(4.2) \quad \tilde{\psi}^{\theta_\star}\langle U_{0:n}\rangle(Z_0) = \tilde{\psi}^\theta\langle U_{0:n}\rangle(Z_0) \quad \mathbb{P}^{\theta_\star}\text{-a.s.}$$

We now show that this implies

$$(H_k) \quad \text{For all } u \in U^k, \text{ we have } \tilde{\psi}^{\theta_\star}\langle (U_{0:(n-k)}, u) \rangle(Z_0) = \tilde{\psi}^\theta\langle (U_{0:(n-k)}, u) \rangle(Z_0) \\ \mathbb{P}^{\theta_\star}\text{-a.s.}$$

by iterative reasoning on $k = 0, \dots, n+1$. First observe that (4.2) corresponds to H_0 (since u is an empty sequence). Now assume that H_k holds for some $k = 0, \dots, n$. Then, for any $v \in U^k$, the set

$$A := \left\{ u \in U : \tilde{\psi}^{\theta_\star}\langle (U_{0:(n-k-1)}, u, v) \rangle(Z_0) = \tilde{\psi}^\theta\langle (U_{0:(n-k)}, u, v) \rangle(Z_0) \right\}$$

has probability 1 under the $\mathbb{P}^{\theta_\star}$ -conditional probability of U_{n-k} given $Z_0, U_{0:(n-k-1)}, V_{(-\infty):(n-k-1)}$. This conditional probability is $\tilde{G}^{\theta_\star}((X_{n-k}, V_{(n-k-r):(n-k-1)}); \cdot)$ defined by (2.7). By (A'-8) and since, given $Z_0, U_{0:(n-k-1)}, V_{(-\infty):(n-k-1)}$, we have $A \in \mathcal{C}^\theta$, $\mathbb{P}^{\theta_\star}$ -a.s., we obtain that H_{k+1} is true. Reasoning by induction, this leads to H_{n+1} , and finally, we get that

$$(H) \quad \text{there exists } z \in Z \text{ such that for all } n \in \mathbb{Z}_{\geq 0} \text{ and all } u \in U^{n+1}, \tilde{\psi}^{\theta_\star}\langle u \rangle(z) = \tilde{\psi}^\theta\langle u \rangle(z).$$

Now let $(z^{(i)}, u_1) \in E^{\theta_\star} \times U$. By definition of E^{θ_\star} , there exists $u \in U^{\mathbb{Z}_{\leq 0}}$, such that $z^{(i)} = \tilde{\Psi}^{\theta_\star}\langle u \rangle$. Now, for all $n \geq p$, we have

$$\begin{aligned} \tilde{\Psi}_{u_1}^\theta \left(\tilde{\Psi}^\theta\langle u_{(-n):0} \rangle(z) \right) &= \tilde{\Psi}^\theta\langle u_{(-n):1} \rangle(z) \\ &= \tilde{\Psi}^{\theta_\star}\langle u_{(-n):1} \rangle(z) \\ &= \tilde{\Psi}_{u_1}^{\theta_\star} \left(\tilde{\Psi}^{\theta_\star}\langle u_{(-n):0} \rangle(z) \right), \end{aligned}$$

where we chose $z \in Z$ in order to apply Assertion (H) in the second equality. On the other hand, by (2.26), we have

$$z^{(i)} = \tilde{\Psi}^{\theta_\star}\langle u \rangle = \lim_{n \rightarrow \infty} \tilde{\Psi}^{\theta_\star}\langle u_{(-n):0} \rangle(z) = \lim_{n \rightarrow \infty} \tilde{\Psi}^\theta\langle u_{(-n):0} \rangle(z),$$

where we again used that z were chosen as in (H) in the last equality. With (A-6) and the previous display we obtain $\tilde{\Psi}_{u_1}^\theta(z^{(i)}) = \tilde{\Psi}_{u_1}^{\theta_\star}(z^{(i)})$. This is true for an arbitrary $(z^{(i)}, u_1) \in E^{\theta_\star} \times U$; hence, we have obtained that the left-hand side of (4.1) is included in its right-hand side.

We now prove the opposite inclusion. Let $\theta_\star, \theta \in \Theta$ such that

$$(4.3) \quad \tilde{\Psi}_u^\theta(z) = \tilde{\Psi}_u^{\theta_\star}(z) \quad \text{for all } (z, u) \in E^{\theta_\star} \times U.$$

Let $v \in U^{\mathbb{Z}_{\leq 0}}$. Take an arbitrary $z^{(i)} \in E^{\theta_\star}$. Then there exists $w \in D^\theta$ such that $z^{(i)} = \tilde{\Psi}^\theta\langle w \rangle$. For all $n \in \mathbb{Z}_{>0}$ and $k = 1, \dots, n$, we get that

$$\tilde{\Psi}^{\theta_\star}\langle v_{(-n+1):(-n+k)} \rangle(z^{(i)}) = \tilde{\Psi}^{\theta_\star}\langle (w, v_{(-n+1):(-n+k)}) \rangle \in E^{\theta_\star}.$$

Applying (4.3) recursively in k , we get that, for any $n \in \mathbb{Z}_{\geq 0}$,

$$\tilde{\Psi}^{\theta_\star}\langle v_{(-n):0} \rangle(z^{(i)}) = \tilde{\Psi}^\theta\langle v_{(-n):0} \rangle(z^{(i)}).$$

Hence, $D^\theta = D^{\theta_\star}$ and by (2.26) and (2.27), we get that for all $v \in D^\theta = D^{\theta_\star}$, $\tilde{\psi}^\theta\langle v \rangle = \tilde{\psi}^{\theta_\star}\langle v \rangle$. By Lemma 8 we have $U_{(-\infty):0} \in D^{\theta_\star}$, $\mathbb{P}^{\theta_\star}$ -a.s., and using (2.31), we get that $\theta \in \langle \theta_\star \rangle$, which concludes the proof. \square

Note that in (4.1), as in the left-hand side of (2.30), the functions $\tilde{\Psi}_u^\theta$ and $\tilde{\Psi}_u^{\theta_\star}$ are only required to coincide on E^{θ_\star} whereas in the left-hand side of (2.16), they coincide on the whole set Z . In some cases, we can prove that the two conditions are the same, so that Proposition 11 and Theorem 12 allow us to conclude that the inclusion in (2.16) is in fact an equality, as in the following result.

Corollary 13. *Suppose that the assumptions of Theorem 12 and (A-7) hold. Let $\theta_\star \in \Theta$. Then the inclusion in (2.30) is an equality. Suppose moreover that θ_\star satisfies the following additional assumption.*

(A-9) For all $\theta \in \Theta$, $u \in U$ and $z \in Z$, if $\tilde{\Psi}_u^\theta$ and $\tilde{\Psi}_u^{\theta_\star}$ coincide on the set $E^{\theta_\star}(z) := \left\{ \tilde{\Psi}^{\theta_\star}(v)(z) : n \in \mathbb{Z}_{\geq 0}, v \in U^n \right\}$, then they also coincide on Z .

Then the inclusion in (2.16) is an equality.

Proof. Applying Proposition 11 and Theorem 12, we get that

$$[\theta_\star] = \left\{ \theta \in [\theta_\star]_G : \tilde{\Psi}_u^\theta(z) = \tilde{\Psi}_u^{\theta_\star}(z) \text{ for all } (z, u) \in E^{\theta_\star} \times U \right\}.$$

Observing that, by (A-7), $\theta \in [\theta_\star]_G$ is equivalent to have $G^\theta = G^{\theta_\star}$, we get that the inclusion in (2.30) is an equality. To prove the second assertion of the corollary, we only need to check that, under (A-9), for all $\theta \in \Theta$ and $u \in U$, if $\tilde{\Psi}_u^\theta$ and $\tilde{\Psi}_u^{\theta_\star}$ coincide on E^{θ_\star} , they must also coincide on Z . It suffices to show that there exists $z \in Z$ such that $E^{\theta_\star}(z) \subset E^{\theta_\star}$. This inclusion is true if $z \in E^{\theta_\star}$ and we conclude by observing that E^{θ_\star} is not empty since it contains Z_1 , $\mathbb{P}^{\theta_\star}$ -a.s., as a consequence of Remark 4. \square

Remark 7. A simple case where (A-9) in Corollary 13 is easy to check is when $p = q = 1$, so that $Z = X$ and $\tilde{\Psi}^{\theta_\star}(v)(z) = \tilde{\psi}_v^{\theta_\star}(z)$ for all $v \in U$ and $z \in X$. See the proof of Theorem 19 for a specific example. However it may happen that (A-9) is not satisfied as will be seen in Remark 8(5). In the linear case, we will characterize $\langle \theta_\star \rangle$ in Lemma 15 without relying on (A-9).

4.2. Vector linear setting. We now consider a $\text{VLODM}(p, q, p', q')$ or a $\text{VLODMX}(p, q, p', q', r)$, that is, we assume the reduced link function to be of the form (2.8). We set in this case $\delta_X(x, x') = |x - x'|$ (resp. $\delta_U(u, u') = |u - u'|$) where $|\cdot|$ denotes an arbitrary norm on $\mathbb{R}^{p'}$ (resp. $\mathbb{R}^{q'}$). The general conditions reduce to the following set of conditions.

(L-1) For all $\theta \in \Theta$, we have that $I_{p'} - \sum_{k=1}^p A_k(\theta)z^k$ is invertible for all $z \in \mathbb{C}$ with $|z| \leq 1$,

where $I_{p'}$ denotes the identity matrix of order p' .

(L-2) For all $\theta \in \Theta$ and $x \in X$, the measure $\tilde{G}^\theta(x; \cdot)$ defined on (U, \mathcal{U}) by (2.6) is non-degenerate in the following sense : there is no affine hyperplane $A \subset \mathbb{R}^{q'}$ such that $\tilde{G}^\theta(x; A) = 1$.

Note that, if $q' = 1$, affine hyperplanes are singletons, hence (L-2) simply means that, for all x and θ , $\tilde{G}^\theta(x; \cdot)$ does not reduce to a unit mass concentrated on a single point. In the case of a VLODMX , we replace (L-2) by the following.

(L'-2) For all $\theta \in \Theta$ and $w \in X \times V^r$, the measure $\tilde{G}^\theta(w; \cdot)$ defined on (U, \mathcal{U}) by (2.7) is non-degenerate in the following sense : there is no affine hyperplane $A \subset \mathbb{R}^{q'}$ such that $\tilde{G}^\theta(w; A) = 1$.

Finally the moment condition (A-5) simplifies in the vector linear case to

(L-3) The invariant probability measure of Definition 3 satisfies, for all $\theta \in \Theta$,

$$(4.4) \quad \mathbb{E}^\theta [\ln^+(|U_0|)] < \infty.$$

We have the following result, whose proof is postponed to [Section A.3](#) for convenience, that relates this set of assumptions to the general ones.

Lemma 14. *Consider the vector linear setting where (2.8) holds, and \mathbf{X} and \mathbf{Y} are closed subset of $\mathbb{R}^{p'}$ and $\mathbb{R}^{q'}$, respectively, with $\delta_{\mathbf{X}}$ and $\delta_{\mathbf{U}}$ being the metrics induced by norms on these spaces. The following assertions hold.*

- (i) Assumption (A-3) holds.
- (ii) Assumption (A-4) is equivalent to (L-1).
- (iii) Assumption (L-3) implies (A-5) for any $x_1^{(i)} \in \mathbf{X}$, and any $u_1^{(i)} \in \mathbf{U}$.
- (iv) Assumption (A-6) holds.
- (v) Assumption (L-2) implies (A-8).
- (vi) Assumption (L'-2) implies (A'-8).

As a consequence, the assumptions of [Theorem 12](#) are implied by (A-1), (L-1), (L-2) and (L-3) in the VLODM case and by (A'-1), (L-1), (L'-2) and (L-3) in the VLODMX case.

We now provides a simple characterization of $\langle \theta_\star \rangle$ in (4.1) in the linear case. The proof of the following result can be found in [Section A.4](#).

Lemma 15. *Suppose that $\{0\} \subsetneq \mathbf{U}$ and that (L-1) holds, and let \mathbf{E}^θ be as in [Definition 7](#). For all $\theta \in \Theta$, define $\mathbf{R}(\cdot; \theta)$ as the rational matrix*

$$(4.5) \quad \mathbf{R}(z; \theta) = \left(\mathbf{I}_{p'} z^p - \sum_{k=1}^p A_k(\theta) z^{p-k} \right)^{-1} \left(\sum_{k=0}^{q-1} B_{k+1}(\theta) z^{q-1-k} \right),$$

which is well defined on $z \in \mathbb{C}$ except for at most finitely many z 's. Then, for all $\theta, \theta_\star \in \Theta$, the two following assertions are equivalent.

- (i) We have $\tilde{\Psi}_u^\theta(z) = \tilde{\Psi}_u^{\theta_\star}(z)$ for all $z \in \mathbf{E}^{\theta_\star}$ and $u \in \mathbf{U}$.
- (ii) The two following identities hold

$$(4.6) \quad \left(\mathbf{I}_{p'} - \sum_{k=1}^p A_k(\theta_\star) \right)^{-1} \omega(\theta_\star) = \left(\mathbf{I}_{p'} - \sum_{k=1}^p A_k(\theta) \right)^{-1} \omega(\theta),$$

$$(4.7) \quad \mathbf{R}(\cdot; \theta_\star) = \mathbf{R}(\cdot; \theta).$$

The identification of a parameter θ based on the equation (4.7) is similar to the identifiability of a vector auto-regressive moving average or order p, q (VARMA(p, q)) model with AR matrices $A_{1:p}(\theta)$ and MA matrices $B_{1:q}(\theta)$. Indeed, in such a model the spectral density matrix takes the form

$$\lambda \mapsto \mathbf{R}(e^{i\lambda}; \theta) \Sigma \mathbf{R}(e^{-i\lambda}; \theta),$$

where Σ is the covariance matrix of the noise. We refer to [\[18\]](#) where identifiable parametrization of ARMA models are discussed. Below we provide an important related result related to this general issue. Let p, q, p', q' be positive integers. For all $A_{1:p} \in (\mathbb{R}^{p' \times p'})^p$ and $B_{1:q} \in (\mathbb{R}^{p' \times q'})^q$, let us define the polynomial matrices respectively valued in $\mathbb{R}^{p' \times p'}$ and $\mathbb{R}^{p' \times q'}$

$$(4.8) \quad \mathbf{P}_p(z; A_{1:p}) = \mathbf{I}_{p'} z^p - \sum_{k=1}^p A_k z^{p-k} \quad \text{and} \quad \mathbf{Q}_q(z; B_{1:q}) = \sum_{k=1}^q B_k z^{q-k}.$$

Note that, for all $A_{1:p} \in (\mathbb{R}^{p' \times p'})^p$, $\mathbf{P}_p(z; A_{1:p})$ in (4.8) must be invertible for $|z|$ large enough (since then $\mathbf{I}_{p'} z^p$ dominates). When a polynomial matrix is invertible for at least one $z \in \mathbb{C}$, then it is invertible for all $z \in \mathbb{C}$, except at most a finite number of them. It is then said to be *non-singular*. Thus, for all $A_{1:p} \in (\mathbb{R}^{p' \times p'})^p$ and $B_{1:q} \in (\mathbb{R}^{p' \times q'})^q$, we can define the rational matrix $\mathbf{P}_p(z; A_{1:p})^{-1} \mathbf{Q}_q(z; B_{1:q})$, which is well defined for all $z \in \mathbb{C}$, except at most a finite number of them.

Lemma 16. *Let p, q, p', q' be positive integers. Then, for any $A_{1:p}^* \in (\mathbb{R}^{p' \times p'})^p$ and $B_{1:q}^* \in (\mathbb{R}^{p' \times q'})^q$, the two following assertion holds.*

- (i) *Suppose that $P_p(\cdot; A_{1:p}^*)$ and $Q_q(\cdot; B_{1:q}^*)$ are left coprime. Then, for all $A_{1:p} \in (\mathbb{R}^{p' \times p'})^p$ and $B_{1:q} \in (\mathbb{R}^{p' \times q'})^q$, we have*

$$P_p(\cdot; A_{1:p}^*)^{-1} Q_q(\cdot; B_{1:q}^*) = P_p(\cdot; A_{1:p})^{-1} Q_q(\cdot; B_{1:q})$$

if and only if $A_{1:p} = A_{1:p}^$ and $B_{1:q} = B_{1:q}^*$.*

- (ii) *Suppose that $P_p(\cdot; A_{1:p}^*)$ and $Q_q(\cdot; B_{1:q}^*)$ are not left coprime. Then, there exist $\tilde{A}_{1:p} \in (\mathbb{R}^{p' \times p'})^p \setminus \{0\}$ and $\tilde{B}_{1:q} \in (\mathbb{R}^{p' \times q'})^q$ such that, for all $\alpha \in \mathbb{R}$, setting $A_{1:p} = A_{1:p}^* + \alpha \tilde{A}_{1:p}$ and $B_{1:q} = B_{1:q}^* + \alpha \tilde{B}_{1:q}$, we have*

$$P_p(\cdot; A_{1:p}^*)^{-1} Q_q(\cdot; B_{1:q}^*) = P_p(\cdot; A_{1:p} + \alpha \tilde{A}_{1:p})^{-1} Q_q(\cdot; B_{1:q}^* + \alpha \tilde{B}_{1:q}) .$$

Two polynomial matrices with the same number p' of rows are said to be left coprime if they admit the identity matrix $I_{p'}$ as a greatest common $p' \times p'$ left divisor (g.c.l.d.), that is, every common left divisor of them is also a left divisor of $I_{p'}$. The set of polynomial matrices of order p' is a non-commutative ring for $p' > 1$. This is why for $p' > 1$ a notion of *left* (or *right*) divisor is necessary. Note, however that if $p' = 1$, saying that $P_p(\cdot; A_{1:p}^*)$ and $Q_q(\cdot; B_{1:q}^*)$ are *left* coprime is equivalent to say that $P_p(\cdot; A_{1:p}^*)$ and the q' row entries of $Q_q(\cdot; B_{1:q}^*)$ (which is q' -dimensional row vector of polynomials of degree at most q) are coprime, that is, they have 1 as greater common divisor. In particular if $p' = q' = 1$, this boils down to say that $P_p(\cdot; A_{1:p}^*)$ and $Q_q(\cdot; B_{1:q}^*)$ have no common roots. The case $p' > 1$ is significantly more complicated and we refer to [22, Chapter III] for an excellent introduction on polynomials on Euclidean rings that applies to matrices of polynomials.

Proof. Let $A_{1:p}^* \in (\mathbb{R}^{p' \times p'})^p$ and $B_{1:q}^* \in (\mathbb{R}^{p' \times q'})^q$. In this proof section, for convenience we denote $P_p(\cdot; A_{1:p}^*)$ and $Q_q(\cdot; B_{1:q}^*)$ by P^* and Q^* .

Proof of Assertion (i). Suppose that P^* and Q^* are left coprime. The Bezout theorem for matrices of polynomials (see e.g. [22, Theorem 3.1]) gives that there exists two polynomial matrices R and S of order $p' \times p'$ and $q' \times p'$ respectively, such that

$$(4.9) \quad I_{p'} = P^* R + Q^* S .$$

Let $A_{1:p} \in (\mathbb{R}^{p' \times p'})^p$ and $B_{1:q} \in (\mathbb{R}^{p' \times q'})^q$, and denote $P = P_p(\cdot; A_{1:p})$ and $Q = Q_q(\cdot; B_{1:q})$. The “if” in Assertion (i) is obvious, so we only need to assume that

$$(4.10) \quad P^{-1} Q = P^{*-1} Q^*$$

and prove that $P = P^*$ (in which case we also get that $Q = Q^*$). Define the rational matrix $U = P P^{*-1}$. Multiplying both sides of (4.9) by U from the left, we have

$$U = P R + P P^{*-1} Q^* S = P R + P P^{-1} Q S = P R + Q S ,$$

where we used (4.10) in the second equality. Hence we get that U is a polynomial matrix and since $U P^* = P$ and both P^* and P are of the form $I_{p'} z^p +$ a polynomial of degree at most $p - 1$, we get that $U = I_{p'}$ and so $P^* = P$, which concludes the proof of (i).

Proof of Assertion (ii). Suppose that P^* and Q^* are not left coprime. Let D be a g.c.l.d. of (P^*, Q^*) . Then D is a polynomial matrix that left-divides P^* and Q^* and changing this polynomial matrix won't change the rational matrix $P^{*-1} Q^*$. The difficulty is to show that we can modify a left divisor of P^* and Q^* in such a way that the resulting P and Q are still of the form (4.8) for some well chosen $A_{1:p} \in (\mathbb{R}^{p' \times p'})^p$ and $B_{1:q} \in (\mathbb{R}^{p' \times q'})^q$. To this end we must first choose D in a special form. Indeed, for any *unimodular* polynomial matrix U , $D U$ is also a g.c.l.d. of (P^*, Q^*) . The polynomial matrix $D U$ is called a right-associate of D , and

by [22, Theorem 22.1], we can choose U so that DU is in *Hermite normal form*, that is such that DU is triangular inferior,

$$DU = \begin{bmatrix} h_{1,1} & 0 & \dots & 0 \\ h_{2,1} & h_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ h_{p',1} & \dots & \dots & h_{p',p'} \end{bmatrix}$$

with the polynomial $h_{i,i}$ is unitary with degree $\deg(h_{i,i})$ strictly larger than those of $h_{j,i}$ for all $j > i$ (that is, the degree on the diagonal dominates those of the same column). Let

$$k = \min \{i = 1, \dots, p' : \deg(h_{i,i}) > 0\} .$$

From what precedes, this min exists (the set is not empty), otherwise we would have $DU = I_{p'}$ and P^* and Q^* would be left coprime. We can thus write, for some $k \in \{1, \dots, p'\}$,

$$DU = \begin{bmatrix} I_{k-1} & 0_{k-1, p'-k+1} \\ 0_{p'-k+1, k-1} & T \end{bmatrix} \quad \text{with} \quad T = \begin{bmatrix} h_{k,k} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ h_{p',k} & \dots & \dots & h_{p',p'} \end{bmatrix} ,$$

where $0_{k,\ell}$ is the zero matrix of size $k \times \ell$. By convention, if $k = 1$, DU reduces to T , that is back to its previous form. The important point is that with this definition of k , we know that $h_{k,k}$ is unitary with $\deg(h_{k,k}) \geq 1$. Now, since DU is a left divisor of P^* and Q^* we may write

$$P^* = (DU)R \quad \text{and} \quad Q^* = (DU)S$$

for some matrices R and S of respective sizes $p' \times p'$ and $p' \times q'$. We write R and S in a block matrix form compatible with that of DU , that is

$$R = \begin{bmatrix} R^{(1)} \\ R^{(2)} \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} S^{(1)} \\ S^{(2)} \end{bmatrix}$$

with $R^{(1)}$ and $S^{(1)}$ of respective sizes $(k-1) \times (k-1)$ and $(k-1) \times q'$. (again with convention that these matrices vanish if $k = 1$). Then we have

$$P^* = \begin{bmatrix} R^{(1)} \\ TR^{(2)} \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} S^{(1)} \\ TS^{(2)} \end{bmatrix}$$

In particular the first row of $TR^{(2)}$ is $h_{k,k}$ times the first row of $R^{(2)}$, and since $h_{k,k}$ is unitary with $\deg(h_{k,k}) \geq 1$, the form of P^* implies that the first row of $R^{(2)}$ is made of polynomials of degrees at most $p-1$ and cannot be zero (since the degree of the (k,k) entry row of P^* is exactly p and all the other entries are zero). Similarly, the first row of $S^{(2)}$ is made of polynomials of degrees at most $q-2$ (since Q^* is of degree $q-1$). Let Δ_k denote the diagonal matrix of order p' with zeros on its diagonal except on the k -th entry where it is 1. Since $\Delta_k R$ and $\Delta_k S$ only keeps the first rows of the block matrices $R^{(2)}$ and $S^{(2)}$, respectively, and put all other entries to zero, we get from what precedes that $\Delta_k R$ and $\Delta_k S$ are of degree at most $p-1$ and $q-1$, respectively, and that $\Delta_k R$ is not zero. Hence we may find $\tilde{A}_{1,p} \in (\mathbb{R}^{p' \times p'})^p \setminus \{0\}$ and $\tilde{B}_{1,q} \in (\mathbb{R}^{p' \times q'})^q$ (with $\tilde{B}_1 = 0$) such that

$$\Delta_k R = Q_p(z; \tilde{A}_{1,p}) \quad \text{and} \quad \Delta_k S = Q_q(z; \tilde{B}_{1,q}) .$$

Then, for all $\alpha \in \mathbb{R}$, setting $A_{1,p} = A_{1,p}^* + \alpha \tilde{A}_{1,p}$ and $B_{1,q} = B_{1,q}^* + \alpha \tilde{B}_{1,q}$, we have

$$P_p(z; A_{1,p}) = P_p(z; A_{1,p}^*) + \alpha \Delta_k R = (DU + \alpha \Delta_k) R ,$$

and, similarly,

$$Q_p(z; B_{1:p}) = Q_q(z; B_{1:q}^*) + \alpha \Delta_k S = (DU + \alpha \Delta_k) S.$$

Then we get

$$\begin{aligned} P_p(\cdot; A_{1:p})^{-1} Q_q(\cdot; B_{1:q}) &= ((DU + \alpha \Delta_k) R)^{-1} (DU + \alpha \Delta_k) S = R^{-1} S \\ &= (DUR)^{-1} DUS \\ &= P^{*-1} Q^*, \end{aligned}$$

which concludes the proof of Assertion (ii). \square

5. APPLICATIONS

We now use our results to derive necessary and sufficient conditions for having identifiability in the examples of Section 3. Many other examples can be achieved by combining various observation kernels and link function with or without exogenous covariates.

5.1. Standard LODMs. Let us apply the results of Section 4.2 in the case of standard LODMs as defined in Definition 9. The moment assumption (L-3) can be readily used for standard LODMs with known or unknown observation kernel, with $|\cdot|$ in (4.4) denoting the usual absolute value. The other assumptions of the general VLODM listed in Section 4.2 can be simplified as follows.

For a standard LODM, Assumption (L-1) becomes

(SL-1) For all $\theta = (\omega, a_{1:p}, b_{1:q})$, we have $1 - \sum_{k=1}^p a_k z^k \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.

In the case of a standard LODM with unknown observation kernel it becomes

(SL'-1) For all $\theta = (\vartheta, \varphi) \in \Theta$ with $\vartheta = (\omega, a_{1:p}, b_{1:q})$, we have $z - \sum_{k=1}^p a_k z^k \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.

As for (L-2), it becomes

(SL-2) For all $x \in X$, $\Upsilon(Y)$ does not degenerate to a single point for $Y \sim G(x, \cdot)$, that is, for all $u \in \mathbb{R}$, we have $G(x, \{\Upsilon(\cdot) = u\}) < 1$;

and, in case of an unknown observation kernel,

(SL'-2) For all $\varphi \in \Phi$, $x \in X$ and $u \in \mathbb{R}$, we have $G^\varphi(x, \{\Upsilon(\cdot) = u\}) < 1$.

Finally (A-7) becomes

(SL-3) For all $x, x' \in X$, $G(x; \cdot) = G(x'; \cdot)$ if and only if $x = x'$;

and, in the case of an unknown observation kernel, it reads as

(SL'-3) For all $\theta = (\vartheta, \varphi)$ and $\theta_* = (\vartheta_*, \varphi_*)$ in Θ , for all $x, x' \in X$, we have

$$G^\varphi(x; \cdot) = G^{\varphi_*}(x'; \cdot) \quad \text{if and only if} \quad \varphi = \varphi_* \quad \text{and} \quad x = x'.$$

This says that the class $[\theta_*]_G$ in (A-7) is given by $[\theta_*]_G = \{\theta = (\vartheta, \varphi) \in \Theta : \varphi = \varphi_*\}$.

Remarkably, all standard LODMs share the same necessary and sufficient condition for identifiability, which can be expressed as follows, using $a_{1:p}^*$ and $b_{1:q}^*$ to denote the true linear coefficients of the linear link function.

(SL-4) The polynomials $z^p - \sum_{k=1}^p a_k^* z^{p-k}$ and $\sum_{k=0}^{q-1} b_{k+1}^* z^{q-1-k}$ have no common complex roots.

We can now state the following result, which says that the true parameter θ_* in the interior of Θ is identifiable if and only if (SL-4) holds.

Theorem 17. Consider a standard LODM(p, q) satisfying (A-1) and (L-3) for some $p, q \in \mathbb{Z}_{>0}$, and suppose that $0 \in U$. In the case of a known observation kernel, suppose that (SL-1)–(SL-3) hold. In the case of an unknown observation kernel, suppose that (SL'-1)–(SL'-3) hold. Then the inclusion in (2.30) is an equality for all $\theta_\star \in \Theta$. In the case of a known observation kernel, Assertions (i) and (ii) below hold for any $\theta_\star = (\omega^\star, a_{1:p}^\star, b_{1:q}^\star) \in \Theta$. In the case of a known observation kernel, Assertions (i) and (iii) below hold for any $\theta_\star = (\omega^\star, a_{1:p}^\star, b_{1:q}^\star, \varphi_\star) \in \Theta$.

- (i) Condition (SL-4) implies that $[\theta_\star]$ reduces to the singleton $\{\theta_\star\}$.
- (ii) If Condition (SL-4) does not hold, then there exists an open segment $I^\star \subset \mathbb{R}^{1+p+q}$ of positive length and containing θ_\star such that $I^\star \cap \Theta \subset [\theta_\star]$.
- (iii) If (SL-4) does not hold, then there exists an open segment $I^\star \subset \mathbb{R}^{1+p+q}$ of positive length and containing $(\omega^\star, a_{1:p}^\star, b_{2:q}^\star)$ such that $\{(\vartheta, \varphi_\star) \in \Theta : \vartheta \in I^\star\} \subset [\theta_\star]$.

Remark 8. Let us briefly comment this result.

- (1) The ergodicity of all the examples of Section 3.1 have been studied in the provided references and the parameter set Θ is always chosen to satisfy the assumptions of Theorem 17 in these references.
- (2) If $p = q = 1$, condition (SL-4) is reduced to $b_1^\star \neq 0$. Let us see what $b_1^\star = 0$ would imply about the identifiability of the model in this simple case. Taking $\tilde{\psi}^\theta$ as in (2.8) with $p = q = p' = q' = 1$, if $\omega(\theta) = \omega$, $A_1(\theta) = a_1^\star$ and $B_1(\theta) = b_1^\star = 0$, then $\{X_k : k \in \mathbb{Z}_{\geq 0}\}$ is a deterministic sequence which, under the stationary distribution, has to be constantly equal to $x^\star = \frac{\omega^\star}{1-a_1^\star}$. But since the distribution of $\{Y_n : n \in \mathbb{Z}\}$ is then uniquely defined by this constant, if one can find a parameter θ with corresponding coefficients ω, a_1, b_1 such that $b_1 = 0$, $(\omega, a_1) \neq (\omega^\star, a_1^\star)$ yielding the same constant $\omega/(1-a_1) = \omega^\star/(1-a_1^\star)$, we see that the model is not identifiable.
- (3) Condition (SL-4) holds for “many” parameters $a_{1:p}^\star, b_{1:q}^\star$, e.g. for Lebesgue almost all ones in \mathbb{R}^{p+q} .
- (4) The identifiability condition (SL-4) is a well known sufficient condition in the standard GARCH(p, q) models, see [14, (A4)] or [2, Condition (2.27)]. Assertion (ii) in Theorem 17 shows that it is also necessary at least for all parameters in the interior set of Θ .
- (5) Suppose that U and X both contain at least two different points and take $a_{1:p}^\star$ and $b_{1:q}^\star$ to be all non-zero. Then it is easy to show for a standard LODM with known observation kernel that $\tilde{\Psi}_u^\theta(z) = \tilde{\Psi}_u^{\theta_\star}(z)$ for all $u \in U$ and $z \in Z$ implies $\theta = \theta_\star$ and thus, we get that the left-hand side of (2.16) reduces to the singleton $\{\theta_\star\}$. Since, as explained previously, (SL-4) is necessary to have $[\theta_\star] = \{\theta_\star\}$ for all θ_\star in the interior set of Θ , we easily get examples for which the inclusion in (2.16) is strict.
- (6) Theorem 17 can be applied to all the models mentioned in Section 3.1. Let us examine the case of the MPINGARCH(p, q) model of [25], which constitutes a rich class of integer valued models. An MPINGARCH(p, q) model is an LODM(p, q) model with unknown observation kernel $G^\varphi(x, \cdot)$ defined as a mixed Poisson distribution with mean x and variance proportional to φ^{-1} , and $\Upsilon(y) = y$. In [25, Theorem 1], the sufficient conditions for having (A-1) imply (L-3) (since $\mathbb{E}^\theta[|U_0|] < \infty$) and (SL-1) (since they imply $\sum_i a_i < 1$, with $a_i \geq 0$). Conditions (SL'-1) and (SL'-2) also hold by definition of $G^\varphi(x, \cdot)$. Hence Theorem 17 applies and we get that (SL-4) is a sufficient condition for identifiability. It is also necessary by Assertion (iii) of the theorem, at least for parameters in the interior of Θ . This condition seems to be missing in [25, Theorem 2].

Proof of Theorem 17. We only consider the case with unknown observation kernel (the case with known observation kernel is obtained by removing the additional parameter φ).

Let $\theta_\star = (\vartheta_\star, \varphi_\star) \in \Theta$ with $\vartheta_\star = (\omega^\star, a_{1:p}^\star, b_{1:q}^\star)$. The assumptions of Theorem 17 are merely adapted from the general VLODM case to the specific parametrization of the standard VLODMs. In particular, we have that (L-1)–(L-3) hold. Hence Lemma 14 implies that (A-3)–(A-8) hold in the general setting with $U = \mathbb{R}$. Applying Theorem 12, we get that (A-2) holds and that

$$\langle \theta_\star \rangle = \left\{ \theta \in \Theta : \tilde{\Psi}_u^\theta(z) = \tilde{\Psi}_u^{\theta_\star}(z) \text{ for all } (z, u) \in E^{\theta_\star} \times U \right\},$$

where $\langle \theta_\star \rangle$ and E^{θ_\star} are as in Definition 6 and Definition 7. Remember that (SL'-3) says that (A-7) holds with

$$[\theta_\star]_G = \{ \theta = (\vartheta, \varphi) \in \Theta : \varphi = \varphi_\star \}.$$

By Corollary 13, we get that the inclusion in (2.30) is an equality and

$$[\theta_\star] = \left\{ \theta = (\vartheta, \varphi_\star) \in \Theta : \tilde{\Psi}_u^\theta(z) = \tilde{\Psi}_u^{\theta_\star}(z) \text{ for all } (z, u) \in E^{\theta_\star} \times U \right\}.$$

Note that we assumed that $0 \in U$ and that (SL'-2) implies $U \neq \{0\}$, hence we can apply Lemma 15 which gives that $[\theta_\star]$ is the set of all $\theta = (\omega, a_{1:p}, b_{1:q}, \varphi_\star) \in \Theta$ such that

$$\begin{aligned} \left(1 - \sum_{k=1}^p a_k \right) \omega^\star &= \left(1 - \sum_{k=1}^p a_k^\star \right) \omega, \\ \text{and } \frac{\sum_{k=0}^{q-1} b_{k+1}^\star z^{q-1-k}}{z^p - \sum_{k=1}^p a_k^\star z^{p-k}} &= \frac{\sum_{k=0}^{q-1} b_{k+1} z^{q-1-k}}{z^p - \sum_{k=1}^p a_k z^{p-k}}. \end{aligned}$$

Applying Lemma 16, we easily get Assertions (i) and (iii). \square

5.2. The bivariate example of Section 3.2. Theorem 17 can be extended to the standard VLODM case of Definition 10. Here, for brevity, we do not re-express the general VLODM assumptions (L-1)–(L-3) in the *standard* setting as we did previously for standard LODMs. We only need to introduce the condition

(SL'-4) The polynomials $z^p - \sum_{k=1}^p A_k^\star z^{p-k}$ and $\sum_{k=0}^{q-1} B_{k+1}^\star z^{q-1-k}$ are left-coprime, which extends (SL-4) to the case $p', q' \geq 1$. The proof of the following result mimics the one of Theorem 17 and is thus omitted.

Theorem 18. *Consider a standard VLODM(p, q, p', q') satisfying (A-1) for some $p, q, p', q' \in \mathbb{Z}_{>0}$. Suppose that $0 \in U$ and that (L-1)–(L-3) and (SL'-3) hold. Then, for all $\theta_\star = (\omega^\star, A_{1:p}^\star, B_{1:q}^\star, \varphi_\star) \in \Theta$, the inclusion in (2.30) is an equality and the two following assertions hold.*

- (i) Condition (SL'-4) implies that $[\theta_\star]$ reduces to the singleton $\{\theta_\star\}$.
- (ii) If (SL'-4) does not hold, then there exists an open segment $I^\star \subset \mathbb{R}^{1+p+q}$ of positive length and containing $(\omega^\star, a_{1:p}^\star, b_{2:q}^\star)$ such that $\{(\vartheta, \varphi_\star) \in \Theta : \vartheta \in I^\star\} \subset [\theta_\star]$.

Remark 9. In Theorem 18, for brevity, we only stated the case with unknown observation kernel. The case with known observation kernel follows by removing the parameters φ and φ_\star in the statement and by replacing (SL'-3) by (SL-3).

Since the bivariate integer valued GARCH model of Section 3.2 is a *standard* VLODM(1, 1, 2, 2) with unknown observation kernel, we just need to check the assumptions of Theorem 18. Ergodicity (hence our Assumption (A-1)) is stated in [9, Theorem 1] under their set of condition (a) on the parameter $\theta = (\omega, A, B, \varphi)$. Their conditions for ergodicity implies some operator norm of A to be strictly less

than 1, which implies $I_2 - A z$ to be invertible for $|z| \leq 1$ and thus (L-1) holds. Since for all $x_1, x_2 > 0$, the bivariate distribution defined by (3.1) has positive probability on all points $(y_1, y_2) \in \mathbb{Z}_{\geq 0}^2$, it cannot have probability one on a line of \mathbb{R}^2 , hence (L-2). Also it is claimed following [9, Theorem 1] that $\mathbb{E}^\theta[U_0]$ is well defined and thus (L-3) holds. Applying Theorem 18, we get that, for any interior point θ_* of the parameter space, Condition (SL'-4) is a necessary and sufficient condition to have identifiability of θ_* . In the bivariate integer valued GARCH model (for which $p = q = 1$), this condition reads for $\theta_* = (\omega^*, A^*, B^*, \varphi^*)$ as $I_2 z - A^*$ and B^* to be left coprime. If θ_* does not satisfy this condition, consistent estimation of θ_* is not possible. Hence we believe that this assumption is missing in [9, Theorem 2]. A precise counter-example is for instance obtained by setting

$$A^* = \alpha \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad B^* = \beta \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

where $\alpha, \beta > 0$ are arbitrary (and chosen in order to make $\theta_* = (\omega^*, A^*, B^*, \varphi^*)$ in the interior of Θ). One can show that $I_2 z - A^*$ and B^* are not left coprime since they both admit the same non-unimodular left divisor $L = \begin{bmatrix} (z - \alpha) & 1 \\ (-\alpha) & 1 \end{bmatrix}$ as shown by the following identities:

$$I_2 z - A^* = L \begin{bmatrix} 1 & -1 \\ 0 & (z - 2\alpha) \end{bmatrix} \quad \text{and} \quad B^* = L \begin{bmatrix} 0 & 0 \\ \beta & \beta \end{bmatrix}.$$

5.3. The non-linear GARCH of Section 3.3. Consider Case 1) of Section 3.3, for which δ is not included in the set of parameters. Then the non-linear GARCH model is a standard VLODM($p, q, 1, 2$) model and identifiability can be treated using Theorem 18. Using that $\Upsilon(y) = ((y^+)^{\delta}, (y^-)^{\delta})$, for all $x > 0$, $\tilde{G}(x, \cdot)$ in (2.6) (we omit θ as G does not depend on θ here) has support included in $\mathbb{R}_{\geq 0} \times \{0\} \cup \{0\} \times \mathbb{R}_{\geq 0}$, where it is defined, for all Borel set $A \subset \mathbb{R}_{\geq 0}$ by

$$\tilde{G}(x; \{0\} \times A) = \mathbb{P}(x(\eta^+)^{\delta} \in A) \quad \text{and} \quad \tilde{G}(x; A \times \{0\}) = \mathbb{P}(x(\eta^-)^{\delta} \in A)$$

It follows that our Assumption (L-2) is equivalent to having that $0 < \mathbb{P}(\eta > 0) < 1$ and that there is no pair $\{u, v\}$, $u \neq v \in \mathbb{R}$, such that $\mathbb{P}(\eta \in \{u, v\}) = 1$, which is exactly the condition appearing in the second part of [17, A3]. Our conditions (L-1) (which here, since $A_k = a_k \geq 0$ for all $k = 1, \dots, p$ simply reads $\sum_k a_k < 1$) and (L-3) are usual byproducts of showing the ergodicity condition (A-1), see [17, Appendix A]. One can thus apply our Theorem 18 (in its know observation kernel version, see Remark 9) and obtain the necessary and sufficient condition (SL'-4) which in the case where the parameters A_k are denoted by $a_k \in \mathbb{R}$ for $k = 1, \dots, p$ and the parameters B_k denoted by $\mathbf{b}_k \in \mathbb{R}^2$ for $k = 1, \dots, q$, becomes, for any $\theta_* = (\omega^*, a_{1:p}^*, \mathbf{b}_{1:q}^*)$ with $\mathbf{b}_k^* = [b_k^*(1) \quad b_k^*(2)]$ for $k = 1, \dots, q$,

(NLG-1) The polynomial $z^p - \sum_{k=1}^p a_k^* z^{p-k}$ have no common complex roots neither with the polynomial $\sum_{k=0}^{q-1} b_{k+1}^*(1) z^{q-1-k}$ nor with the polynomial $\sum_{k=0}^{q-1} b_{k+1}^*(2) z^{q-1-k}$.

This condition is similar to that appearing in the identifiability condition [17, A4] used in a mis-specified context. Our result shows that this condition is necessary in the interior of the parameter set in the well-specified case, and remains valid for much larger choices of observation kernels.

5.4. The SETPAR model of Section 3.4. We have the following result for the self-excited threshold Poisson autoregression model.

Theorem 19. *Consider the SETPAR model introduced in Section 3.4. Let*

$$\Theta \subset \{(\omega_1, \omega_2, a_1, a_2, b_1, b_2, r) \in \mathbb{R}_{>0}^2 \times [0, 1)^2 \times \mathbb{R}_{\geq 0} \times [0, 1) \times \mathbb{Z}_{\geq 0} : a_2 + b_2 < 1\},$$

Then (A-1) holds. Let $\theta_\star = (\omega_1^\star, \omega_2^\star, a_1^\star, a_2^\star, b_1^\star, b_2^\star, r^\star) \in \Theta$ satisfy at least one of the two following conditions.

- (i) $b_1^\star > 0$ and $r^\star \geq 1$;
- (ii) $b_2^\star > 0$.

Then we have

$$(5.1) \quad [\theta_\star] = \{\theta \in \Theta : \psi_y^\theta(x) = \psi_y^{\theta_\star}(x) \text{ for all } x \in \mathbb{R} \text{ and } y \in \mathbb{Z}_{\geq 0}\},$$

where $\psi_y^\theta(x)$ is defined by (3.2).

Remark 10. Let us briefly comment this result.

- (1) The case where neither (i) nor (ii) hold ($b_1^\star = b_2^\star = 0$ or $r^\star = b_2^\star = 0$) is somehow degenerate, similarly to the non-threshold case mentioned in Remark 8(2). We think it should be treated separately but we omit this very special case here for brevity.
- (2) As explained after (2.16), the identity (5.1) is the best we could hope for this model since the distribution $\tilde{\mathbb{P}}^\theta$ of the observations is entirely determined by the mapping $(u, x) \mapsto \psi_u^\theta(x)$ on $\mathbb{Z}_{\geq 0} \times \mathbb{R}_{\geq 0}$.
- (3) The identity (5.1) shows in particular that θ_\star is not identifiable if $r^\star = 0$ (since changing b_1^\star will have no effect on the mapping $(u, x) \mapsto \psi_u^\theta(x)$ on $\mathbb{Z}_{\geq 0} \times \mathbb{R}_{\geq 0}$). Another case of non-identifiability is when $a_1^\star = a_2^\star$ and $\omega_1^\star + b_1^\star(r+1) = \omega_2^\star + b_2^\star(r+1)$. In such a case, we have for all $x \in \mathbb{R}$,

$$\omega_1^\star + a_1^\star x + b_1^\star y = \omega_2^\star + a_2^\star x + b_2^\star y \quad \text{at } y = r+1.$$

Then, setting $\theta = (\omega_1^\star, \omega_2^\star, a_1^\star, a_2^\star, b_1^\star, b_2^\star, r^\star + 1)$, we immediately have that $\psi_y^\theta(x) = \psi_y^{\theta_\star}(x)$ for all $x \in \mathbb{R}$ and $y \in \mathbb{Z}_{\geq 0}$. In particular consistent estimation of θ_\star as claimed in [28, Theorem 2] is not possible for such a parameter θ_\star .

Proof of Theorem 19. A natural choice for \mathbf{X} is $\mathbb{R}_{>0}$ but in order to meet Assumption (A-3) with $\delta_{\mathbf{X}}(x, x') = |x - x'|$ we take $\mathbf{X} = \mathbb{R}_{\geq 0}$ with $G^\theta(0, \cdot)$ arbitrarily set to be Bernoulli with mean 1/2 for convenience (it actually has no influence on \mathbb{P}^θ since $\omega_0, \omega_1 > 0$ in the definition of Θ). We set $\Upsilon(y) = y$ so that $U_k = Y_k$ for all k and the reduced link function $\tilde{\psi}$ is the same as the non-reduced one. Moreover since $p = q = 1$, we are in the case where $Z_k = X_k$ for all k , $\mathbb{Z} = \mathbf{X} = \mathbb{R}_{\geq 0}$ and $\tilde{\Psi}_u^\theta = \tilde{\psi}_u^\theta$ for all u . By [28, Theorem 1], with Θ satisfying the given condition, Assumption (A-1) holds, and, moreover, for any $\ell > 0$ and $\theta \in \Theta$, $\mathbb{E}^\theta[U_0^\ell] < \infty$ (in fact, one can prove that, for any $\theta \in \Theta$, there exists $\ell > 0$ such that $\mathbb{E}^\theta[\exp(\ell U_0)] < \infty$). This moment condition implies that the log moment condition (A-5) holds for any $x_1^{(i)} \in \mathbf{X}$. Clearly, we have $\text{Lip}_1^\theta = a_1 \vee a_2$ and, since $p = q = 1$, $\text{Lip}_n^\theta \leq (\text{Lip}_1^\theta)^n$ for all $n \geq 1$. Thus the above condition on Θ also implies (A-4). As for (A-6), it trivially holds (since $y = u$ is fixed in this condition). Hence with Lemma 8, we get that (A-1)–(A-6) holds, with definitions (2.26) for checking (A-2). Assumption (A-7) is also immediate with $[\theta]_G = \Theta$ and Proposition 11 gives that, for any $\theta_\star \in \Theta$,

$$[\theta_\star] = \langle \theta_\star \rangle,$$

where $\langle \theta_\star \rangle$ can be defined by (2.31). Assumption (A-8) holds by Remark 6(1). Hence all the assumptions of Theorem 12 hold. Take now $\theta_\star = (\omega_1^\star, \omega_2^\star, a_1^\star, a_2^\star, b_1^\star, b_2^\star, r^\star) \in \Theta$ satisfying (i) or (ii). Let $\theta \in \Theta$, $u \in \mathbf{U}$ and $z \in \mathbf{Z}$. To conclude the proof, it is now sufficient to check that if $\tilde{\Psi}_u^\theta$ and $\tilde{\Psi}_u^{\theta_\star}$ coincide on $\{\tilde{\Psi}^{\theta_\star}\langle v \rangle(z) : n \in \mathbb{Z}_{\geq 0}, v \in \mathbf{U}^n\}$, then they must coincide on \mathbf{Z} , so that we can apply Corollary 13. Observe that by definition of the link function in (3.2), if (i) or (ii) holds, then $v \mapsto \psi_v^{\theta_\star}(z)$ takes at least two different values on $\mathbf{U} = \mathbb{Z}_{\geq 0}$. Since $\tilde{\Psi}^{\theta_\star}\langle v \rangle(z) = \psi_v^{\theta_\star}(z)$ for all

$v \in U$, these two different values belong to $\{\tilde{\Psi}^{\theta_*} \langle v \rangle(z) : n \in \mathbb{Z}_{\geq 0}, v \in U^n\}$. Now, since $\tilde{\Psi}_u^{\theta_*} = \psi_u^{\theta_*}$ and $\tilde{\Psi}_u^\theta = \psi_u^\theta$ are affine functions, if they coincide in two points they must coincide everywhere, and the proof is concluded. \square

5.5. The PARX model of Section 3.5. We have the following result for the Poisson autoregression model with exogenous covariates.

Theorem 20. *Consider the PARX model defined in Section 3.5, which is a VLDMX($p, q, 1, 1 + d$). Suppose that (A'-1), (L-1) and (L-3) hold, and that the exogenous kernel H satisfies the following.*

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

Then, for all $\theta_ \in \Theta$, the equivalent class $[\theta_*]$ reduces to the singleton $\{\theta_*\}$.*

Remark 11. Let us briefly comment this result.

- (1) (P-1) is a natural assumption as it basically says that the covariates $f_1(V_k), \dots, f_d(V_k)$ are not linearly related conditionally to V_{k-1} . If they were, it would suggest using a smaller set of covariates.
- (2) The ergodicity of PARX models (our assumption (A'-1) is treated in [1, Theorem 1] under some assumption on the covariate kernel H (in their assumption 2). Their assumption 3 used for proving (A'-1) implies $\sum_k a_k < 1$ which implies our assumption (L-1) (since $a_k \geq 0$ for all $k = 1, \dots, p$). Note also that [1, Theorem 1] implies that $\mathbb{E}^\theta[|U_0|] < \infty$ and thus our assumption (L-3). On the other hand, their identifiability condition [1, Assumption 5] include a condition on parameters $a_{1:p}^*, b_{1:q}^*$ similar to our condition (SL-4) used for standard LODM's in Theorem 17 above. Theorem 20 shows that such a condition can in fact be dropped in case of exogenous covariates provided that the mild condition (P-1) holds.
- (3) This result is similar to Theorem 17 for the standard LODM. It is of interest to note that there is no additional condition on $\gamma_{1:d}^*$ for identifiability.
- (4) Theorem 20 easily extends to more general observation kernels (known or unknown) G^θ , provided that, as in Theorem 17, assumptions (SL-2)-(SL-3) hold if the observation kernel is known, or (SL'-2)-(SL'-3) if it is unknown. However, the ergodicity would require a specific treatment in these cases, as only the Poisson case has been considered up to our knowledge.

Proof of Theorem 20. For the PARX model we set $\Upsilon(y, v) = (y, f_1(v), \dots, f_d(v)) \in U = \mathbb{R}^{1+d}$ and $X = \mathbb{R}_{\geq 0}$ (with $G(0, \cdot)$ arbitrarily set, say, to be Bernoulli with mean $1/2$, as in the proof of Theorem 19) and (A-3) holds with $\delta_X(x, x') = |x - x'|$ and $\delta_U(u, u') = |u - u'|$ where $|\cdot|$ here is an arbitrary norm on \mathbb{R}^{1+d} . In this case, we have that, for all $(x, v) \in X \times V$, if $(Y, W) \sim \tilde{G}((x, v), \cdot)$ with Y valued in \mathbb{R} , W valued in \mathbb{R}^d and \tilde{G} defined by (2.7), we have that Y and W are independent and Y follows a Poisson distribution. Thus (L'-2) is equivalent to (P-1). We can thus apply Lemma 14 and get that Assumptions (A-3), (A-4), (A-5), (A-6) and (A'-8) hold with $U = \mathbb{R}^{1+d}$. Then, having assumed (A'-1), we can apply Theorem 12 and get that (A-2) holds as well as the identity (4.1). Assumption (A-7) is immediate with $[\theta_*]_G = \Theta$ and, Applying Proposition 11 and the previous display we get that

$$[\theta_*] = \left\{ \theta \in \Theta : \tilde{\Psi}_u^\theta(z) = \tilde{\Psi}_u^{\theta_*}(z) \text{ for all } (z, u) \in E^{\theta_*} \times U \right\}.$$

where $\langle \theta_* \rangle$ and E^{θ_*} are as in Definition 6 and Definition 7. Lemma 15 and the definitions of $A_{1:p}$ and $B_{1:q}$ now give that $[\theta_*]$ is the set of all $\theta = (\omega, a_{1:p}, b_{1:q}, \gamma_{1:d}) \in$

Θ such that

$$\begin{aligned} \left(1 - \sum_{k=1}^p a_k\right) \omega^* &= \left(1 - \sum_{k=1}^p a_k^*\right) \omega, \\ \frac{\sum_{k=0}^{q-1} b_{k+1}^* z^{q-1-k}}{z^p - \sum_{k=1}^p a_k^* z^{p-k}} &= \frac{\sum_{k=0}^{q-1} b_{k+1} z^{q-1-k}}{z^p - \sum_{k=1}^p a_k z^{p-k}}, \\ \left(z^p - \sum_{k=1}^p a_k^* z^{p-k}\right)^{-1} d_i^* z^{q-1} &= \left(z^p - \sum_{k=1}^p a_k z^{p-k}\right)^{-1} d_i z^{q-1} \quad \text{for } 1 \leq i \leq d. \end{aligned}$$

Note that the last line in this set of equations is equivalent to

$$a_k^* = a_k \quad \text{and} \quad d_i = d_i^* \quad \text{for all } 1 \leq i \leq d \quad \text{and} \quad 1 \leq k \leq p.$$

But the two first lines of the previous display give that $\omega = \omega^*$ and $b_{1,q}^* = b_{1,q}$, thus $\theta = \theta_*$ and the proof is concluded. \square

ACKNOWLEDGMENTS

In the first version of this contribution, we neither investigated the VLODM case nor the case with exogenous covariates. We are grateful to the two referees that reviewed the first submission, whose fruitful and constructive comments motivated these demanding extensions.

APPENDIX A. POSTPONED PROOFS

A.1. Proof of Lemma 8. We first derive the following result.

Lemma 21. (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 $n \in \mathbb{Z}_{>0}$.

Proof. By (2.22), (2.23) and (2.15), we have, for all $n \in \mathbb{Z}_{>0}$, using the convention $\text{Lip}_m^\theta = 1$ for $m \leq 0$,

$$(A.1) \quad \sup_{u \in \mathbb{U}^n, v \in \mathbb{Z}^2} \frac{\delta_Z \circ \tilde{\Psi}^\theta \langle u \rangle^{\otimes 2}(v)}{\delta_Z(v)} \leq \mathbb{1}_{\{n < q\}} \vee \left(\max_{0 \leq j < p} \text{Lip}_{n-j}^\theta \right).$$

Hence (A-4) implies that there exists $m \geq 1$ and $L \in (0, 1)$ such that, for all $u \in \mathbb{U}^{m+1}$, $\tilde{\Psi}^\theta \langle u \rangle$ is L -Lipschitz. Now observe that, by (2.14), for all $n = km + r$ with $k \geq 0$ and $0 \leq r < m$, for all $u = u_{-n:0} \in \mathbb{U}^{n+1}$, we can write $\tilde{\Psi}^\theta \langle u \rangle$ as

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

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

$$\delta_X(\tilde{\psi}^\theta \langle u \rangle(z), \tilde{\psi}^\theta \langle u \rangle(z')) \leq \delta_Z(\tilde{\Psi}^\theta \langle u \rangle(z), \tilde{\Psi}^\theta \langle u \rangle(z')) \leq L' L^k \delta_Z(z, z').$$

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

We can now prove Lemma 8. Let $\theta, \theta_* \in \Theta$ and let $z^{(i)} \in \mathbb{Z}$. Denote for all $n \in \mathbb{Z}_{\geq 0}$,

$$X^{(n)} = \tilde{\psi}^\theta \langle U_{(-n):0} \rangle(z^{(i)}).$$

Then, by (2.13) we have, for all $n \in \mathbb{Z}_{>0}$,

$$X^{(n)} = \tilde{\psi}^\theta \langle U_{-n+1:0} \rangle \circ \tilde{\Psi}_{U_{-n}}^\theta(z^{(i)}),$$

and, by (2.22), we get

$$(A.2) \quad \delta_X(X^{(n)}, X^{(n-1)}) \leq \text{Lip}_n^\theta \delta_Z(z^{(i)}, \tilde{\Psi}_{U_{-n}}^\theta(z^{(i)})).$$

Using (2.23) with $z^{(i)} = (x^{(i)}, u^{(i)})$ and $x_1^{(i)} = \dots = x_p^{(i)}$ and $u_1^{(i)} = \dots = u_{q-1}^{(i)}$ we get that

$$\delta_Z \left(z^{(i)}, \tilde{\Psi}_{U_{-n}}^\theta(z^{(i)}) \right) = \delta_X \left(x_1^{(i)}, \tilde{\psi}_{(u^{(i)}, U_{-n})}^\theta(x^{(i)}) \right) \bigvee \delta_U(u_1^{(i)}, U_{-n})$$

Hence, for all $\alpha > 0$, Condition (2.24) implies as $n \rightarrow \infty$,

$$\delta_Z \left(z^{(i)}, \tilde{\Psi}_{U_{-n}}^\theta(z^{(i)}) \right) = O(e^{\alpha n}) \quad \tilde{\mathbb{P}}^{\theta*}\text{-a.s.}$$

The last display with (A.2) and Lemma 21 gives that, $\tilde{\mathbb{P}}^{\theta*}\text{-a.s.}$, $\{X^{(n)} : n \in \mathbb{Z}_{\geq 0}\}$ is a Cauchy sequence, hence converges in \mathbf{X} . Therefore, $U_{(-\infty):0} \in D^\theta$, $\tilde{\mathbb{P}}^{\theta*}\text{-a.s.}$. By (2.1) or (2.2) depending whether an ODM or an ODMX is considered, we also have, under \mathbb{P}^θ , for all $n \in \mathbb{Z}_{\geq 0}$, $X_1 = \tilde{\psi}^\theta \langle U_{-n:0} \rangle(Z_{-n})$. Thus, (2.22) also implies

$$\delta_X(X_1, X^{(n)}) \leq \text{Lip}_{n+1}^\theta \delta_Z \left(Z_{-n}, z^{(i)} \right) \quad \mathbb{P}^\theta\text{-a.s.}$$

By stationarity, $\delta_Z(Z_{-n}, z^{(i)})$ is bounded in probability under \mathbb{P}^θ , hence $X^{(n)}$ converges to X_1 in probability if (A-4) holds. We thus obtain (2.17), and since this holds for all $\theta \in \Theta$, Assumption (A-2) holds.

Let us now check (2.30). Take $\theta_*, \theta \in \Theta$ and suppose that θ belongs to the set in left-hand side of the inclusion (2.30), that is, $G^\theta = G^{\theta*}$ and $\tilde{\Psi}_u^\theta(z) = \tilde{\Psi}_u^{\theta*}(z)$ for all $(z, u) \in E^{\theta*} \times U$. By Remark 4 we have $Z_k \in E^{\theta*}$, $\mathbb{P}^{\theta*}\text{-a.s.}$. Thus we get that $\tilde{\Psi}_{U_k}^\theta(Z_k) = \tilde{\Psi}_{U_k}^{\theta*}(Z_k)$ $\mathbb{P}^{\theta*}\text{-a.s.}$ and with Remark 2, we obtain that, $\mathbb{P}^{\theta*}\text{-a.s.}$, (Y_k, X_k) (resp. (Y_k, X_k, V_k)) satisfy the iterative equations (2.1) (resp. (2.2)), and by (A-1) (resp. (A'-1)), we conclude that $\mathbb{P}^\theta = \mathbb{P}^{\theta*}$. Hence $\theta \in [\theta_*]$ and (2.30) is proved.

Finally, we check that (2.21) holds when $\tilde{\psi}_u^\theta$ is continuous for all $u \in U^q$. Since we have shown that $U_{(-\infty):0} \in D^\theta$, $\tilde{\mathbb{P}}^{\theta*}\text{-a.s.}$ and $\tilde{\mathbb{P}}^{\theta*}$ is shift invariant, we have, for all $k \in \mathbb{Z}$,

$$\tilde{\psi}^\theta \langle U_{(-\infty):k} \rangle = \lim_{n \rightarrow \infty} \tilde{\psi}^\theta \langle U_{-n:k} \rangle(z^{(i)}) \quad \tilde{\mathbb{P}}^{\theta*}\text{-a.s.}$$

Observe that, for all $n \geq p \vee q$, we have, for all $u_{(-n):0} \in U^{n+1}$,

$$\tilde{\psi}^\theta \langle u_{(-n):0} \rangle(z^{(i)}) = \tilde{\psi}_{u_{(-q+1):0}}^\theta \left(\left(\tilde{\psi}^\theta \langle u_{(-n):(-j)} \rangle(z^{(i)}) \right)_{-p \leq j \leq -1} \right).$$

By continuity of $\tilde{\psi}_u^\theta$ and using the previous display, we can take the limit as $n \rightarrow \infty$ under $\tilde{\mathbb{P}}^{\theta*}$ and obtain (2.21).

A.2. Proof of Proposition 11. First observe that (2.17) implies for all $\theta \in \Theta$,

$$(A.3) \quad \tilde{\mathbb{P}}^\theta [Y_1 \in \cdot | Y_{(-\infty):0}] = G^\theta \left(\tilde{\psi}^\theta \langle U_{(-\infty):0} \rangle; \cdot \right) \quad \tilde{\mathbb{P}}^{\theta*}\text{-a.s.}$$

Let us now show that any $\theta \in [\theta_*]$ belongs to $[\theta_*]_G \cap \langle \theta_* \rangle$, that is, $\theta \in [\theta_*]_G$, and (2.20) and (2.21) hold true. Since $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta*}$, (A.3), which also holds with θ replaced by θ_* , yields

$$G^\theta \left(\tilde{\psi}^\theta \langle U_{(-\infty):0} \rangle; \cdot \right) = G^{\theta*} \left(\tilde{\psi}^{\theta*} \langle U_{(-\infty):0} \rangle; \cdot \right) \quad \tilde{\mathbb{P}}^{\theta*}\text{-a.s.}$$

By (A-7), we obtain that $\theta \in [\theta_*]_G$ and (2.20) holds. By Lemma 5, (2.17) implies (2.18), and using $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta*}$, we obtain (2.21). Thus $\theta \in \langle \theta_* \rangle$.

It remains to show that $[\theta_*]_G \cap \langle \theta_* \rangle \subseteq [\theta_*]$. We prove this inclusion in the case of an ODMX satisfying (A'-1). (The case of an ODM is readily obtained by removing the variables V_k 's in the reasoning). Let $\theta \in [\theta_*]_G$ such that (2.20) and (2.21) hold true. Since (2.17) holds with θ replaced by θ_* , (2.20) gives that $X_1 = \tilde{\psi}^\theta \langle U_{(-\infty):0} \rangle$ $\mathbb{P}^{\theta*}\text{-a.s.}$ Since $\mathbb{P}^{\theta*}$ is shift invariant, we get, for all $k \in \mathbb{Z}$, $X_{k+1} = \tilde{\psi}^\theta \langle U_{(-\infty):k} \rangle$

$\mathbb{P}^{\theta_\star}$ -a.s. With (2.21), we obtain $X_1 = \tilde{\psi}_{U_{(-q+1):0}}^\theta (X_{(-p+1):0})$ $\mathbb{P}^{\theta_\star}$ -a.s. Since $\mathbb{P}^{\theta_\star}$ is shift invariant, we thus have, for all $k \in \mathbb{Z}$,

$$(A.4) \quad X_{k+1} = \tilde{\psi}_{U_{(k+1-q):k}}^\theta (X_{(k+1-p):k}) \quad \mathbb{P}^{\theta_\star}\text{-a.s.}$$

On the other hand, by definition of $\mathbb{P}^{\theta_\star}$ and using (A-7) with $\theta \in [\theta_\star]_G$, we have that

$$\mathbb{P}^{\theta_\star} [Y_1 \in \cdot \mid X_{(-\infty):1}, Y_{(-\infty):0}, V_{(-\infty):0}] = G^{\theta_\star}(X_1; \cdot) = G^\theta(X_1; \cdot) \quad \mathbb{P}^{\theta_\star}\text{-a.s.}$$

And using again that $\mathbb{P}^{\theta_\star}$ is shift-invariant, for all $k \in \mathbb{Z}$,

$$\mathbb{P}^{\theta_\star} [Y_k \in \cdot \mid X_{(-\infty):k}, Y_{(-\infty):(k-1)}, V_{(-\infty):(k-1)}] = G^\theta(X_k; \cdot) \quad \mathbb{P}^{\theta_\star}\text{-a.s.}$$

This, with (A.4), shows that $\mathbb{P}^{\theta_\star}$ is a shift-invariant solution of (2.2). By (A'-1), we conclude that $\mathbb{P}^{\theta_\star} = \mathbb{P}^\theta$, and thus $\theta \in [\theta_\star]$.

A.3. Proof of Lemma 14. Assertion (i) is obvious.

Proof of Assertion (ii). In the vector linear setting we have, for all $n \in \mathbb{Z}_{>0}$ and all $(z, z', u) \in \mathbb{Z}^2 \times U^n$,

$$\delta_X(\tilde{\psi}^\theta \langle u \rangle(z), \tilde{\psi}^\theta \langle u \rangle(z')) = |\tilde{\psi}_n^\theta(z - z')| ,$$

where $|\cdot|$ is a norm on $\mathbb{R}^{p'}$ and, for all $n \in \mathbb{Z}_{>0}$, $\tilde{\psi}_n^\theta$ is a linear mapping from $Z = \mathbb{R}^{p' * p + q' * (q-1)}$ to $X = \mathbb{R}^{p'}$ recursively defined by setting, for all $w = w_{1:(p+q+1)} \in Z = (\mathbb{R}^{p'})^p \times (\mathbb{R}^{q'})^{q-1}$, $\tilde{\psi}_n^\theta(w) = x_n$ with

$$(A.5) \quad \begin{cases} u_j = w_{p+q+j} , & -q < j \leq -1 , \\ u_j = 0 , & 0 \leq j \leq n-1 , \\ x_j = w_{p+j} , & -p < j \leq 0 , \\ x_j = \sum_{k=1}^p A_k(\theta) x_{j-k} + \sum_{k=1}^q B_k(\theta) u_{j-k} , & 1 \leq j . \end{cases}$$

In particular, we have, for all $j \geq q$,

$$x_j = \sum_{k=1}^p A_k(\theta) x_{j-k} ,$$

and this equation is also true for $j = 1, \dots, q-1$ if the last $q-1$, $\mathbb{R}^{q'}$ -valued, component of w are equal to zero. It is well known that the Lipschitz norm of such iterative linear functions goes to zero if and only if (L-1) holds.

Proof of Assertion (iii). Take an arbitrary $x_1^{(i)} \in X$. If $q > 1$, take also an arbitrary $u_1^{(i)} \in U$ and set $u^{(i)} = (u_1^{(i)}, \dots, u_1^{(i)}) \in U^{q-1}$. Then, since $\tilde{\psi}_u^\theta(x)$ is of the form (2.8), there exists constants $C_1, C_2 > 0$ only depending on $\theta, x_1^{(i)}$ and $u_1^{(i)}$ such that, for all $u \in U$,

$$\delta_X \left(x_1^{(i)}, \tilde{\psi}_{(u^{(i)}, u)}^\theta(x^{(i)}) \right) \leq C_1 + C_2 |u| .$$

Assertion (iii) follows.

Assertion (iv) is obvious.

Proof of Assertions (v) and (vi). See Remark 6 (2) in the case of a VLDM. The case of a VLDMX is similar.

A.4. Proof of Lemma 15. Note that, by Lemma 14 (ii), in the vector linear case, the set E^θ of Definition 7 is well defined under (L-1). We need the following result whose proof is straightforward, and thus omitted.

Lemma 22. Suppose that (L-1) holds and let $\theta \in \Theta$. Let $\ell^1(\mathbb{Z}, \mathbb{U})$ denote the set of sequences in $\mathbb{U}^{\mathbb{Z}}$ that are absolutely summable. For any $u \in \ell^1(\mathbb{Z}, \mathbb{U})$, there is a unique $x \in \ell^\infty(\mathbb{Z}, \mathbb{X})$ (the set of bounded sequences valued in \mathbb{X}) such that

$$(A.6) \quad x_j = \omega(\theta) + \sum_{k=1}^p A_k(\theta) x_{j-k} + \sum_{k=1}^q B_k(\theta) u_{j-i}, \quad j \in \mathbb{Z}.$$

This unique solution is given by

$$x_j = \left(I_{p'} - \sum_{k=1}^p A_k(\theta) \right)^{-1} \omega(\theta) + \int_{-\pi}^{\pi} e^{i\lambda(j+p-q)} \mathbf{R}(e^{i\lambda}; \theta) \hat{u}(\lambda) d\lambda, \quad j \in \mathbb{Z},$$

where \mathbf{R} is defined by (4.5) and \hat{u} denotes the Fourier series of u defined by

$$\hat{u}(\lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} u_k e^{-i\lambda k}, \quad \lambda \in \mathbb{R}.$$

Let D^θ and E^θ be as in (2.25) and Definition 7. Then D^θ contains $\ell^1(\mathbb{Z}, \mathbb{U})$ and, for any $u \in \ell^1(\mathbb{Z}, \mathbb{U})$, defining x as the unique solution of (A.6) in $\ell^\infty(\mathbb{Z}, \mathbb{X})$, we have, for all $t \in \mathbb{Z}$,

$$(x_{(t-p+1):t}, u_{(t-q+1):(t-1)}) = \tilde{\Psi}^\theta \langle u_{(-\infty):(t-1)} \rangle \in E^\theta.$$

We can now prove Lemma 15.

Proof of Lemma 15. Step 1: Assertion (i) implies Assertion (ii). Let $\theta, \theta_\star \in \Theta$ satisfying Assertion (i) and let us show that (4.6) and (4.7) hold. Take any $u \in \ell^1(\mathbb{Z}, \mathbb{U})$ and $n \in \mathbb{Z}_{\geq 0}$. By Lemma 22, $u \in D^{\theta_\star}$ and we have

$$\begin{aligned} \tilde{\Psi}^{\theta_\star} \langle u_{(-\infty):0} \rangle &= \tilde{\Psi}^{\theta_\star} \langle u_{(-n):0} \rangle (\tilde{\Psi}^{\theta_\star} \langle u_{(-\infty):(-n-1)} \rangle) \\ &= \tilde{\Psi}^\theta \langle u_{(-n):0} \rangle (\tilde{\Psi}^{\theta_\star} \langle u_{(-\infty):(-n-1)} \rangle), \end{aligned}$$

where the second equality follows from applying successively Assertion (i) with $u = u_k$ and $z = \tilde{\Psi}^{\theta_\star} \langle u_{(-\infty):(k-1)} \rangle$ for $k = -n, -n+1, \dots, 0$. On the other hand by definition of Lip_n^θ in (2.22), we have, setting $z_n^\star := \tilde{\Psi}^{\theta_\star} \langle u_{(-\infty):(-n-1)} \rangle$ and $z_n = \tilde{\Psi}^\theta \langle u_{(-\infty):(-n-1)} \rangle$,

$$\begin{aligned} \delta_Z \left(\tilde{\Psi}^\theta \langle u_{(-\infty):0} \rangle, \tilde{\Psi}^\theta \langle u_{(-n):0} \rangle (z_n^\star) \right) &= \delta_Z \left(\tilde{\Psi}^\theta \langle u_{(-n):0} \rangle (z_n), \tilde{\Psi}^\theta \langle u_{(-n):0} \rangle (z_n^\star) \right) \\ &\leq \text{Lip}_n^\theta \delta_Z(z_n, z_n^\star). \end{aligned}$$

Since $u \in \ell^1(\mathbb{Z}, \mathbb{U})$, we have that (z_n) and (z_n^\star) are summable sequences (as a consequence of Lemma 22) and so $\delta_Z(z_n, z_n^\star)$ is bounded as $n \rightarrow \infty$. Using Lemma 14 (ii) and we conclude that the upper bound in the last display converges to 0 as $n \rightarrow \infty$. By definition of z_n^\star and the previous display, this gives that $\tilde{\Psi}^\theta \langle u_{(-\infty):0} \rangle = \tilde{\Psi}^{\theta_\star} \langle u_{(-\infty):0} \rangle$. Shifting the sequence u , we also have that $\tilde{\Psi}^\theta \langle u_{(-\infty):t} \rangle = \tilde{\Psi}^{\theta_\star} \langle u_{(-\infty):t} \rangle$ for all $t \in \mathbb{Z}$ and by Lemma 22, this implies that θ and θ_\star share the same unique solution $x \in \ell^\infty(\mathbb{Z}, \mathbb{X})$ to the equation (A.6). Using the explicit form of this solution in the same lemma, we get that, for all $u \in \ell^1(\mathbb{Z}, \mathbb{U})$ and $\tau \in \mathbb{Z}$,

$$\left(I_{p'} - \sum_{k=1}^p A_k(\theta) \right)^{-1} \omega(\theta) + \alpha_\tau(u; \theta) = \left(I_{p'} - \sum_{k=1}^p A_k(\theta_\star) \right)^{-1} \omega(\theta_\star) + \alpha_\tau(u; \theta_\star),$$

where, for all $\theta \in \Theta$, $u \in \ell^1(\mathbb{Z}, \mathbb{U})$ and $\tau \in \mathbb{Z}$, we set $\alpha_\tau(u; \theta) = \int_{-\pi}^{\pi} e^{i\lambda\tau} \mathbf{R}(e^{i\lambda}; \theta) \hat{u}(\lambda) d\lambda$. Since we assumed $\{0\} \subsetneq \mathbb{U}$, we can successively take u as the zero sequence ($u_k = 0$ for all k , implying $\hat{u} \equiv 0$) or proportional to the

impulse sequence ($u_0 \neq 0$, $u_k = 0$ for all $k \neq 0$, implying $\hat{u} \equiv u_0/(2\pi)$), the previous display successively leads to (4.6) and

$$\int_{-\pi}^{\pi} e^{i\lambda\tau} \mathbf{R}(e^{i\lambda}; \theta) d\lambda = \int_{-\pi}^{\pi} e^{i\lambda\tau} \mathbf{R}(e^{i\lambda}; \theta_*) , \quad \text{for all } \tau \in \mathbb{Z} ,$$

which implies (4.7).

Step 2: Assertion (ii) implies Assertion (i). Let $\theta, \theta_* \in \Theta$ satisfying (4.6) and (4.7), and let us show that Assertion (ii) holds. First take $u \in \ell^1(\mathbb{Z}, \mathbb{U})$. By Lemma 22, (4.6) and (4.7) imply that the recursive equation (A.6) and the one with θ replaced by θ_* share the same bounded solution. Moreover, we have $u \in D^{\theta_*} \cap D^{\theta}$ and since $\tilde{\Psi}^{\theta} \langle u_{(-\infty):0} \rangle$ and $\tilde{\Psi}^{\theta_*} \langle u_{(-\infty):0} \rangle$ are given by the same solution they are equal. Hence we obtain that

$$\tilde{\Psi}_u^{\theta}(z) = \tilde{\Psi}_u^{\theta_*}(z) \quad \text{for all } (z, u) \in E_1^{\theta_*} \times \mathbb{U} ,$$

where $E_1^{\theta_*} = \left\{ \tilde{\Psi}^{\theta_*} \langle v \rangle : v \in \ell^1(\mathbb{Z}_{\leq 0}, \mathbb{U}) \right\}$. To get Assertion (i), since $z \mapsto \tilde{\Psi}_u^{\theta'}(z)$ is continuous for $\theta' = \theta, \theta_*$ and for any $u \in \mathbb{R}^{q'}$, it is now sufficient to prove that $E_1^{\theta_*}$ is dense in E^{θ_*} . To this end, pick $z \in E^{\theta_*}$. Then there exists $u \in D^{\theta_*}$ such that

$$(A.7) \quad z = \tilde{\Psi}^{\theta} \langle u \rangle .$$

Define, for any $n \in \mathbb{Z}_{\geq 0}$, we introduce the truncated sequence

$$v_k^{(n)} = \begin{cases} u_k & \text{if } k \in \{-n, \dots, 0\} \\ 0 & \text{otherwise.} \end{cases}$$

Then $v^{(n)} \in \ell^1(\mathbb{Z}, \mathbb{U})$ and we have

$$z_n := \tilde{\Psi}^{\theta_*} \langle v_{(-\infty):0}^{(n)} \rangle \in E_1^{\theta_*} .$$

Moreover, we can write, denoting by $0_{(-\infty):0}$ the null sequence in $\mathbb{U}^{\mathbb{Z}_{\leq 0}}$,

$$z_n = \tilde{\Psi}^{\theta_*} \langle v_{(-n):0} \rangle \left(\tilde{\Psi}^{\theta_*} \langle 0_{(-\infty):0} \rangle \right) = \tilde{\Psi}^{\theta_*} \langle u_{(-n):0} \rangle \left(\tilde{\Psi}^{\theta_*} \langle 0_{(-\infty):0} \rangle \right) .$$

By (2.26) and (A.7) we thus have $z = \lim_{n \rightarrow \infty} z_n$, and since z is arbitrary in E^{θ_*} we have shown that $E_1^{\theta_*}$ is dense in E^{θ_*} and the proof is concluded. \square

REFERENCES

- [1] Arianna Agosto, Giuseppe Cavaliere, Dennis Kristensen, and Anders Rahbek. Modeling corporate defaults: Poisson autoregressions with exogenous covariates (parx). *Journal of Empirical Finance*, 38:640 – 663, 2016. ISSN 0927-5398. doi: <https://doi.org/10.1016/j.jempfin.2016.02.007>. Recent developments in financial econometrics and empirical finance.
- [2] István Berkes, Lajos Horváth, and Piotr Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, 9(2):201–227, 2003. ISSN 1350-7265. doi: <https://doi.org/10.3150/bj/1068128975>. URL <https://doi.org/10.3150/bj/1068128975>.
- [3] Krishnan Bhaskaran, Antonio Gasparrini, Shakoor Hajat, Liam Smeeth, and Ben Armstrong. Time series regression studies in environmental epidemiology. *International journal of epidemiology*, page dyt092, 2013.
- [4] Tim Bollerslev. Glossary to arch (garch). Technical report, CREATES Research Paper, September 2008.
- [5] P. Bougerol and N. Picard. Stationarity of garch processes and of some nonnegative time series. *J. Econometrics*, 52(1992):115 – 127, 1992. ISSN 0304-4076. doi: [10.1016/0304-4076\(92\)90067-2](https://doi.org/10.1016/0304-4076(92)90067-2).

- [6] Vasiliki Christou and Konstantinos Fokianos. Estimation and testing linearity for non-linear mixed Poisson autoregressions. *Electron. J. Stat.*, 9(1):1357–1377, 2015. doi: 10.1214/15-EJS1044. URL <https://doi.org/10.1214/15-EJS1044>.
- [7] Vasiliki Christou and Konstantinos Fokianos. On count time series prediction. *Journal of Statistical Computation and Simulation*, 85(2):357–373, 2015.
- [8] DR Cox. Statistical analysis of time-series: some recent developments. *Scand. J. Statist.*, 8(2):93–115, 1981. ISSN 0303-6898.
- [9] Yan Cui and Fukang Zhu. A new bivariate integer-valued garch model allowing for negative cross-correlation. *TEST*, 27(2):428–452, Jun 2018. ISSN 1863-8260. doi: 10.1007/s11749-017-0552-4. URL <https://doi.org/10.1007/s11749-017-0552-4>.
- [10] Richard A Davis, WTM Dunsmuir, and Y Wang. Modeling time series of count data. *Statistics textbooks and monographs*, 158:63–114, 1999.
- [11] Randal Douc, François Roueff, and Tepmony Sim. Handy sufficient conditions for the convergence of the maximum likelihood estimator in observation-driven models. *Lithuanian Mathematical Journal*, 55(3):367–392, 2015. doi: 10.1007/s10986-015-9286-8.
- [12] René Ferland, Alain Latour, and Driss Oraichi. Integer-valued GARCH process. *J. Time Ser. Anal.*, 27(6):923–942, 2006. ISSN 0143-9782. doi: 10.1111/j.1467-9892.2006.00496.x.
- [13] K. Fokianos and D. Tjøstheim. Log-linear poisson autoregression. *J. of Multivariate Analysis*, 102(3):563–578, 2011.
- [14] Christian Francq and Jean-Michel Zakoian. Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4):605–637, 2004.
- [15] Christian Francq and Jean-Michel Zakoian. A tour in the asymptotic theory of garch estimation. *Handbook of Financial Time Series*, pages 85–111, 2009.
- [16] Christian Francq and Jean-Michel Zakoian. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2011.
- [17] Tawfik Hamadeh and Jean-Michel Zakoian. Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *J. Statist. Plann. Inference*, 141(1):488–507, 2011. ISSN 0378-3758. doi: 10.1016/j.jspi.2010.06.026. URL <https://doi.org/10.1016/j.jspi.2010.06.026>.
- [18] E. J. Hannan and Manfred Deistler. *The statistical theory of linear systems*, volume 70 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2012. ISBN 978-1-611972-18-4. doi: 10.1137/1.9781611972191.ch1. URL <https://doi.org/10.1137/1.9781611972191.ch1>. Reprint of the 1988 original [MR0940698].
- [19] AR Ives, B Dennis, KL Cottingham, and SR Carpenter. Estimating community stability and ecological interactions from time-series data. *Ecological monographs*, 73(2):301–330, 2003.
- [20] Roman Liesenfeld and Jean-Francois Richard. Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of empirical finance*, 10(4):505–531, 2003.
- [21] Alexander M Lindner. Stationarity, mixing, distributional properties and moments of garch (p, q)-processes. In *Handbook of financial time series*, pages 43–69. Springer, 2009.
- [22] C. C. MacDuffee. *The Theory of Matrices*. Springer-Verlag, 1933.
- [23] Robert S Pindyck and Daniel L Rubinfeld. *Econometric models and economic forecasts*, volume 4. Irwin/McGraw-Hill Boston, 1998.

- [24] Tina Hviid Rydberg and Neil Shephard. Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics*, 1(1):2–25, 2003.
- [25] Rodrigo B. Silva and Wagner Barreto-Souza. Flexible and robust mixed poisson ingarch models. *Journal of Time Series Analysis*, 40(5):788–814, 2019.
- [26] Tepmony Sim. *Maximum likelihood estimation in partially observed Markov models with applications to time series of counts*. Theses, Télécom ParisTech, March 2016. URL <https://pastel.archives-ouvertes.fr/tel-01458087>.
- [27] Tepmony Sim, Randal Douc, and François Roueff. General-order observation-driven models: ergodicity and consistency of the maximum likelihood estimator. preprint, April 2019. URL <https://hal.archives-ouvertes.fr/hal-01383554>.
- [28] Chao Wang, Heng Liu, Jian-Feng Yao, Richard A Davis, and Wai Keung Li. Self-excited threshold poisson autoregression. *Journal of the American Statistical Association*, 109(506):777–787, 2014.
- [29] Scott L Zeger. A regression model for time series of counts. *Biometrika*, 75(4):621–629, 1988.
- [30] Fukang Zhu. A negative binomial integer-valued GARCH model. *J. Time Series Anal.*, 32(1):54–67, 2011. ISSN 0143-9782. doi: 10.1111/j.1467-9892.2010.00684.x. URL <http://dx.doi.org/10.1111/j.1467-9892.2010.00684.x>.
- [31] Fukang Zhu. Modeling time series of counts with COM-Poisson INGARCH models. *Math. Comput. Modelling*, 56(9-10):191–203, 2012. ISSN 0895-7177. doi: 10.1016/j.mcm.2011.11.069. URL <https://doi.org/10.1016/j.mcm.2011.11.069>.
- [32] Fukang Zhu. Zero-inflated Poisson and negative binomial integer-valued GARCH models. *J. Statist. Plann. Inference*, 142(4):826–839, 2012. ISSN 0378-3758. doi: 10.1016/j.jspi.2011.10.002. URL <https://doi.org/10.1016/j.jspi.2011.10.002>.

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

LTCI, TÉLÉCOM PARIS, INSTITUT POLYTECHNIQUE DE PARIS, 19 PLACE MARGUERITE PEREY,,
 91120 PALAISEAU, FRANCE
E-mail address: roueff@telecom-paristech.fr

MIT RESEARCH UNIT, INSTITUTE OF TECHNOLOGY OF CAMBODIA, 12156 PHNOM PENH, CAMBODIA
E-mail address: tepmony.sim@itc.edu.kh