



HAL
open science

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. 2019. hal-02088860v1

HAL Id: hal-02088860

<https://hal.science/hal-02088860v1>

Preprint submitted on 4 Apr 2019 (v1), last revised 8 Sep 2020 (v3)

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

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

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 GARCH(p, q) model, a simple sufficient condition has been established in [1] 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 log-linear Poisson GARCH or the NBIN-GARCH models.

1. INTRODUCTION

Observation-driven models were introduced in [6] and have received considerable attention since. They are commonly used for modeling various non-linear times series in applications ranging from economics (see [18]), environmental study (see [2]), epidemiology and public health study (see [24, 7, 9]), finance (see [16, 19, 10, 13]) and population dynamics (see [14]).

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 GARCH(p, q) model, for which an explicit sufficient condition appears in [1], their condition (2.27). We will in fact prove that this condition is necessary and sufficient for the identifiability and extends to a much larger class of observation-driven models than the GARCH(p, q) model. See [Theorem 8](#) below and the comments following this result.

Let us now formally introduce the class of observation-driven models. Throughout the paper we use the notation $u_{\ell:m} := (u_{\ell}, \dots, u_m)$ for $\ell \leq m$, with the convention that $u_{\ell:m}$ is the empty sequence if $\ell > m$, so that, for instance $(x_{0:(-1)}, y) = y$. The observation-driven time series model can formally be defined as follows.

Definition 1 (Observation-driven model (ODM)). Let (X, \mathcal{X}) and (Y, \mathcal{Y}) be two measurable spaces, called the *latent space* and the *observation space*, and let (Θ, Δ) be a compact metric space, called the *parameter space*. Let $\{(x_{1:p}, y_{1:q}) \mapsto \psi_{y_{1:q}}^{\theta}(x_{1:p}) : \theta \in \Theta\}$ be a family of measurable functions from $(X^p \times Y^q, \mathcal{X}^{\otimes p} \otimes \mathcal{Y}^{\otimes q})$ to (X, \mathcal{X}) , called the *link functions* and let $\{G^{\theta} : \theta \in \Theta\}$ be a family of probability kernels on $X \times Y$, called the *observation kernels*. A time series $\{Y_k : k \geq -q + 1\}$ valued in Y is said to be distributed according to an *observation-driven model of order (p, q)* (hereafter, ODM(p, q)) with link function

Date: April 4, 2019.

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

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

ψ^θ and observation kernel G^θ if there exists a process $\{X_k : k \geq -p+1\}$ on $(\mathsf{X}, \mathcal{X})$ such that for all $k \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}$,

$$(1.1) \quad \begin{aligned} Y_k | \mathcal{F}_k &\sim G^\theta(X_k; \cdot), \\ X_{k+1} &= \psi_{Y_{k-q+1:k}}^\theta(X_{(k-p+1):k}), \end{aligned}$$

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

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

$$(1.2) \quad g^\theta(x; y) > 0,$$

and also, to avoid a trivial degenerate case, that ν is non-degenerate, that is, its support contains at least two points.

One of the most popular examples in this class of models is the general GARCH(p, q) model introduced by [3], where $\mathsf{X} = \mathbb{R}_+$, $\mathsf{Y} = \mathbb{R}$, $G^\theta(x; \cdot)$ is (usually but not necessarily) the centered Gaussian distribution of variance x and the link function ψ^θ given by

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

with $\theta = (\omega, a_{1:p}, b_{1:q})$ with $\omega > 0$ and $a_{1:p}, b_{1:q} \geq 0$. This model was extensively studied, see, for example, [5, 11, 12, 17, 13] and the references therein. Many other examples have been derived from this class, see [4]. The fact that the function ψ^θ is linear with respect to the x_i 's and $\Upsilon(y_i)$'s for some mapping Υ (here $\Upsilon(y) = y^2$), as well as the fact that $G^\theta(x; \cdot)$ is a distribution parameterized that does not depend on θ and for which x governs a scale parameter are features that are often shared among these conditionally heteroscedastic models. If $G^\theta(x; \cdot)$ is a discrete distribution, this latter property is not satisfied, which seriously complicates the theoretical analysis of such models, as explained in [23]. Let us recall an example of such a model, namely, the log-linear Poisson GARCH(p, q), see [10].

Example 1. The Log-linear Poisson GARCH(p, q) Model is a LODM(p, q) parameterized by $\theta = (\omega, a_{1:p}, b_{1:q}) \in \Theta \subset \mathbb{R}^{p+q+1}$ with observations space $\mathsf{Y} = \mathbb{Z}_+$ and hidden variables space $\mathsf{X} = \mathbb{R}$, link function given by

$$\psi_y^\theta(x) = \omega + \sum_{i=1}^p a_i x_i + \sum_{i=1}^q b_i \log(1 + y_i),$$

and $G(x, \cdot)$ is the Poisson distribution with mean e^x , that is, ν is the counting measure on \mathbb{Z}_+ and $g^\theta(x; y) = e^x y^{-e^x} / (y!)$.

Many other examples of count time series have been proposed. It is for instance possible to let the parameter θ contain a sub-parameter φ involved in the conditional distribution $G^\theta(x; \cdot)$. It is the case in our next example introduced in [25].

Example 2. The NBIN-GARCH(p, q) model is a LODM(p, q) parameterized by $\theta = (\omega, a_{1:p}, b_{1:q}, r) \in \Theta \subset \mathbb{R}_+^* \times \mathbb{R}_+^{p+q} \times \mathbb{R}_+^*$ with observations space $\mathcal{Y} = \mathbb{Z}_+$ and hidden variables space $\mathcal{X} = \mathbb{R}_+$, link function given by

$$\psi_y^\theta(x) = \omega + \sum_{i=1}^p a_i x_i + \sum_{i=1}^q b_i y_i,$$

and $G(x, \cdot)$ is the the negative binomial distribution with shape parameter $r > 0$ and mean $r x$, that is, ν is the counting measure on \mathbb{Z}_+ and

$$(1.3) \quad g^\theta(x; y) = \frac{\Gamma(r+y)}{y! \Gamma(r)} \left(\frac{1}{1+x} \right)^r \left(\frac{x}{1+x} \right)^y.$$

The rest of the paper is organized as follows. [Section 2](#) contains additional notation and definitions that will be used throughout the paper. Our main results can be found in [Section 3](#), some proofs of which are postponed to [Section 4](#).

2. PRELIMINARIES

2.1. Admissible mappings and linearly observation driven models. We introduce the following definition in order to simplify the form of link function.

Definition 2 (Admissible mapping). Consider an ODM(p, q) with link function ψ^θ . Let Υ be a measurable function from $(\mathcal{Y}, \mathcal{Y})$ to a measurable space $(\mathcal{U}, \mathcal{U})$. If there exists $\tilde{\psi}^\theta : \mathcal{X}^p \times \mathcal{U}^q \rightarrow \mathcal{X}$ measurable, called the *reduced link function* such that, for all $(x, y_{1:q}) \in \mathcal{X}^p \times \mathcal{Y}^q$,

$$(2.1) \quad \psi_{y_{1:q}}^\theta(x) = \tilde{\psi}_{u_{1:q}}^\theta(x) \quad \text{with} \quad u_k = \Upsilon(y_k) \quad \text{for} \quad 1 \leq k \leq q,$$

then we call Υ an *admissible mapping* and it follows that, setting $U_k = \Upsilon(Y_k)$ for all $k > -q$, $\{U_k : k > -q\}$ is an ODM(p, q) with link function $\tilde{\psi}^\theta$ and observation kernel $\tilde{G}(x, \cdot) = G(x, \Upsilon^{-1}(\cdot))$ on the observation space $(\mathcal{U}, \mathcal{U})$. Moreover, for all $k \geq 0$, the conditional distribution of (Y_k, X_{k+1}) given \mathcal{F}_k only depends on

$$(2.2) \quad Z_k = (X_{(k-p+1):k}, U_{(k-q+1):(k-1)}) \in \mathcal{Z},$$

where we defined

$$(2.3) \quad \mathcal{Z} = \mathcal{X}^p \times \mathcal{U}^{q-1} \quad \text{endowed with the } \sigma\text{-field } \mathcal{Z} = \mathcal{X}^{\otimes p} \otimes \mathcal{U}^{\otimes(q-1)}.$$

A subclass of ODM of order (p, q) consists in those having an admissible mapping Υ whose corresponding reduced mapping $\tilde{\psi}_u^\theta(x)$ is linear in (u, x) , that is, for all $x = x_{1:p} \in \mathcal{X}^p$, $u = u_{1:q} \in \mathcal{U}^q$,

$$(2.4) \quad \tilde{\psi}_u^\theta(x) = \omega + \sum_{i=1}^p a_i x_i + \sum_{i=1}^q b_i u_i,$$

In particular the parameters $(\omega, a_{1:p}, b_{1:q})$ must constitute a part of the parameter θ (possibly whole of it). We will denote this part by ϑ and the remaining part (if any) of θ by φ . We call such a model a *linearly observation-driven model* of order (p, q) . The formal definition is as follows.

Definition 3 (Linearly observation-driven model (LODM)). Consider an ODM of order (p, q) . It is called a *linearly observation-driven model of order (p, q)* (hereafter, LODM(p, q)) if

- (i) All $\theta \in \Theta$ can be written as $\theta = (\vartheta, \varphi)$ with $\vartheta \in \mathbb{R}^{1+p+q}$.

- (ii) The latent space X is a closed subset of \mathbb{R} , and, for all $x = x_{1:p} \in \mathsf{X}^p$, $y = y_{1:q} \in \mathsf{Y}^q$, and $\theta = (\vartheta, \varphi) \in \Theta$ with $\vartheta = (\omega, a_{1:p}, b_{1:q})$,

$$(2.5) \quad \psi_y^\theta(x) = \omega + \sum_{i=1}^p a_i x_i + \sum_{i=1}^q b_i \Upsilon(y_i) ,$$

where $\Upsilon : \mathsf{Y} \rightarrow \mathbb{R}$ is measurable with $\nu \circ \Upsilon^{-1}$ being non-degenerate.

Remark 1. The standard GARCH(p, q) model is a special case of LODM(p, q), in which case $\mathsf{X} = \mathbb{R}_+$, $\mathsf{Y} = \mathbb{R}$, $\Upsilon(y) = y^2$, and $G^\theta(x, \cdot)$ is a centered distribution with variance x , most commonly the normal distribution. Similarly the log-linear Poisson Garch model of [Example 1](#) is an LODM(p, q) by taking $\Upsilon(y) = \ln(1 + y)$. Obviously, the NBIN-GARCH model of [Example 2](#) is also an LODM(p, q). Among these three special cases, only the NBIN-GARCH model has an observation kernel G^θ depending on a subpart of the parameter θ through r , which appears in the right-hand side of (1.3). Hence this r corresponds to the φ appearing in the general setting of [Definition 3](#).

2.2. Iterations of the link function. We now introduce iterated versions of the link function ψ^θ . Let Υ be an admissible mapping as given by [Definition 2](#), with corresponding reduced link function $\tilde{\psi}^\theta$, and let Z be defined by (2.3). In this context, throughout the paper, for all $j \in \{1, \dots, p + q - 1\}$, we denote by $\Pi_j(z)$ the j -th entry of $z \in \mathsf{Z}$.

We define for any $y_{0:k} \in \mathsf{Y}^{k+1}$ the mapping $\psi^\theta \langle y_{0:k} \rangle : \mathsf{Z} \rightarrow \mathsf{X}$ through a set of recursive equations of order (p, q) . Namely, for all $n \in \mathbb{N}$, $y_{0:k} \in \mathsf{Y}^{k+1}$ and $z \in \mathsf{Z}$,

$$(2.6) \quad \psi^\theta \langle y_{0:k} \rangle (z) := x_{k+1} ,$$

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

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

In this set of equation the last line is applied recursively so that in fact, for all $j \geq 0$, x_{j+1} only depends on z and $y_{0:j}$.

3. MAIN RESULTS

To study the identifiability of such models 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 satisfying (1.1).

This ergodic property is the cornerstone for making statistical inference theory work and we provide simple general conditions in [8] for $p = q = 1$ and in [20, 21, Chapter 5] 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 4 (Stationary distributions \mathbb{P}^θ and $\tilde{\mathbb{P}}^\theta$). Suppose that (A-1) holds. We define the distributions \mathbb{P}^θ and $\tilde{\mathbb{P}}^\theta$ as follows.

- a) \mathbb{P}^θ denotes the distribution on $(\mathsf{X} \times \mathsf{Y})^{\mathbb{Z}}, (\mathcal{X} \times \mathcal{Y})^{\otimes \mathbb{Z}}$ of the stationary solution of (1.1) extended on $k \in \mathbb{Z}$, with $\mathcal{F}_k = \sigma(X_{-\infty:k}, Y_{-\infty:(k-1)})$;

b) $\tilde{\mathbb{P}}^\theta$ denotes the projection of \mathbb{P}^θ on the component $\mathcal{Y}^{\mathbb{Z}}$.

We also use the symbols \mathbb{E}^θ and $\tilde{\mathbb{E}}^\theta$ to denote the expectations corresponding to \mathbb{P}^θ and $\tilde{\mathbb{P}}^\theta$, respectively. We further denote by π_X^θ and π_Y^θ the marginal distributions of X_0 and Y_0 under \mathbb{P}^θ , on (X, \mathcal{X}) and (Y, \mathcal{Y}) , respectively. Moreover, for all $\theta, \theta' \in \Theta$, we write $\theta \sim \theta'$ if and only if $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta'}$. This defines an equivalence relation on the parameter set Θ and the corresponding equivalence class of θ is denoted by $[\theta] := \{\theta' \in \Theta : \theta \sim \theta'\}$.

In this context, identifiable parameters are defined as all the parameters θ_* for which $[\theta_*]$ is reduced to the singleton $\{\theta_*\}$. Without identifiability, the consistency of any estimator of θ_* is not possible. Nevertheless a weaker notion of consistency, the *equivalence-class consistency* introduced by [15] can still be established and the consistency in the usual sense is then equivalent to θ_* being identifiable.

As a byproduct of the proof of (A-1), one usually obtains a function $V_X : X \rightarrow \mathbb{R}_+$ of interest, common to all $\theta \in \Theta$, such that, for all $\theta \in \Theta$, $\pi_X^\theta(V_X) < \infty$, see [20, Chapter 5]. It turns out that, since $\pi_Y^\theta = \pi_X^\theta G^\theta$, such a condition allows one to check moment conditions on π_Y^θ such as Conditions (3.8) and (3.11) below.

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-2) We can write $\theta_* = (\vartheta_*, \varphi_*)$ and, for all $\theta = (\vartheta, \varphi)$ in Θ and $x, x' \in X$,

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

The “if” in Assumption (A-2) simply means that G^θ does not depend on the component ϑ of the parameter, so could be written as G^φ . The “only if” part then says that, adopting temporarily this notation, $(\varphi, x) \mapsto G^\varphi(x, \cdot)$ is one-to-one. In many examples G^θ does not depend on θ at all, in which case the component φ can be discarded. It is the case for the Log-linear Poisson GARCH model of Example 1 but not for the NBIN-GARCH model of Example 2, where $\varphi = r$.

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

Proposition 5. *Let p and q be two positive integers. Consider an ODM(p, q) with admissible mapping Υ and let $\theta_* \in \Theta$. Assume that (A-1) and (A-2) hold. Suppose moreover that, for all $\theta \in \Theta$, there exists a measurable function $\psi^\theta(\cdot) : \mathcal{Y}^{\mathbb{Z}^-} \rightarrow X$ such that*

$$(3.1) \quad X_1 = \psi^\theta \langle Y_{-\infty:0} \rangle \quad \mathbb{P}^\theta\text{-a.s.}$$

Then the equivalent class $[\theta_*]$ of Definition 4 coincides with the set $\langle \theta_* \rangle$ defined as the set of all $\theta = (\vartheta, \varphi_*) \in \Theta$ satisfying the two following equations

$$(3.2) \quad \psi^\theta \langle Y_{-\infty:0} \rangle = \psi^{\theta_*} \langle Y_{-\infty:0} \rangle \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.},$$

$$(3.3) \quad \psi^\theta \langle Y_{-\infty:0} \rangle = \psi_{Y_{(-q+1):0}}^{\theta_*} \left(\left(\psi^{\theta_*} \langle Y_{-\infty:j} \rangle \right)_{-p \leq j \leq -1} \right) \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

The proof is postponed to Section 4.1 for convenience. For the moment, let us provide important insights for $\psi^\theta \langle Y_{-\infty:0} \rangle$ appearing in Proposition 5. These insights are threefold. First we show in Lemma 6 that (3.1) can be verified using $\tilde{\mathbb{P}}^\theta$ only. Second, in Lemma 7, that (3.1) is implied by some Lipschitz condition on the iterates of the link function ψ^θ and a moment condition on π_Y . Finally, we examine the special case of LODM(p, q) for which the link function is linear, leading to Theorem 8 below, that will conclude this section.

Lemma 6. *Consider an ODM(p, q) for some $p, q \geq 1$ and assume that (A-1) holds true. Let $\theta \in \Theta$ and consider a measurable function $\psi^\theta \langle \cdot \rangle : \mathcal{Y}^{\mathbb{Z}^-} \rightarrow \mathcal{X}$. Then (3.1) is satisfied if and only if the two following equations hold.*

$$(3.4) \quad \psi^\theta \langle Y_{-\infty:0} \rangle = \psi_{Y_{-q+1:0}}^\theta \left((\psi^\theta \langle Y_{-\infty:j} \rangle)_{-p \leq j \leq -1} \right) \quad \tilde{\mathbb{P}}^\theta\text{-a.s.}$$

$$(3.5) \quad \tilde{\mathbb{P}}^\theta [Y_1 \in \cdot | Y_{-\infty:0}] = G^\theta (\psi^\theta \langle Y_{-\infty:0} \rangle, \cdot) \quad \tilde{\mathbb{P}}^\theta\text{-a.s.}$$

Proof. Suppose that (3.1) holds true. Since \mathbb{P}^θ is shift invariant, it can be extended to all time instants $k \in \mathbb{Z}$,

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

But then (3.4) and (3.5) follows from the model equations (1.1).

Suppose now that (3.4) and (3.5) hold true. Since \mathbb{P}^θ is shift invariant, they are extended to all time instants $k \in \mathbb{Z}$ in the form

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

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

Defining $X'_k = \psi^\theta \langle Y_{-\infty:(k-1)} \rangle$ 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 (1.1). By uniqueness of \mathbb{P}^θ assumed in (A-1), we get that $\{(X'_k, Y_k) : k \in \mathbb{Z}\}$ has distribution \mathbb{P}^θ and so (3.1) holds since it is true with X'_1 replacing X_1 , by the mere definition of X'_1 . \square

Whenever we need some metric on the space \mathcal{Z} , we assume the following.

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

Recall that, for any finite \mathcal{Y} -valued sequence y , the mapping $\psi^\theta \langle y \rangle$ is defined by (2.6) following the recursion in (2.7). Define, for all $n \in \mathbb{Z}_+^*$, the Lipschitz constant for $\psi^\theta \langle y \rangle$, uniform over $y \in \mathcal{Y}^n$,

$$(3.6) \quad \text{Lip}_n^\theta = \sup \left\{ \frac{\delta_{\mathcal{X}}(\psi^\theta \langle y \rangle(z), \psi^\theta \langle y \rangle(z'))}{\delta_{\mathcal{Z}}(z, z')} : (z, z', y) \in \mathcal{Z}^2 \times \mathcal{Y}^n \right\},$$

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

$$(3.7) \quad \delta_{\mathcal{Z}}(v) = \left(\max_{1 \leq k \leq p} \delta_{\mathcal{X}} \circ \Pi_k^{\otimes 2}(v) \right) \vee \left(\max_{p < k < p+q} \delta_{\mathcal{U}} \circ \Pi_k^{\otimes 2}(v) \right).$$

Consider the following assumption on the link function.

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

We have the following result, whose proof is postponed to Section 4.2 for convenience.

Lemma 7. *Let p and q be two positive integers. Consider an ODM(p, q) with admissible mapping Υ . Assume that (A-1), (A-3) and (A-4) hold, and take $\theta, \theta_* \in \Theta$. Suppose that there exists $x_1^{(i)} \in \mathcal{X}$ and $y_1^{(i)} \in \mathcal{Y}$ such that the constant vectors $x^{(i)} = (x_1^{(i)}, \dots, x_1^{(i)}) \in \mathcal{X}^p$ and $u^{(i)} = (\Upsilon(y_1^{(i)}), \dots, \Upsilon(y_1^{(i)})) \in \Upsilon(\mathcal{Y})^{q-1}$ satisfy*

$$(3.8) \quad \int \phi^{(i)} d\pi_{\mathcal{Y}}^\theta \quad \text{and} \quad \int \phi^{(i)} d\pi_{\mathcal{Y}}^{\theta_*} \quad \text{are finite,}$$

where, for all $y \in \mathcal{Y}$,

$$\phi^{(i)}(y) = \ln^+ \left(\delta_{\mathcal{X}} \left(x_1^{(i)}, \psi^\theta \langle y \rangle((x^{(i)}, u^{(i)})) \right) \vee \delta_{\mathcal{U}}(\Upsilon(y_1^{(i)}), \Upsilon(y)) \right)$$

with the convention $(x^{(i)}, u^{(i)}) = x^{(i)}$ and $\delta_U(\Upsilon(y_1^{(i)}), \Upsilon(y)) = 0$ if $q = 1$. Then

$$(3.9) \quad \psi^\theta \langle Y_{-\infty:0} \rangle = \lim_{n \rightarrow \infty} \psi^\theta \langle Y_{-n:0} \rangle((x^{(i)}, u^{(i)}))$$

is well defined $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. and $\tilde{\mathbb{P}}^\theta$ -a.s. and satisfies (3.1). If moreover, $x \mapsto \psi_y^\theta(x)$ is continuous for all $y \in \mathcal{Y}^q$, then (3.3) holds.

We now examine the application of Proposition 5 for an LODM(p, q), as given by Definition 3. In this case, the recursion (2.7) is defined with $\tilde{\psi}^\theta$ given by (2.4). Start this recursion with two different z and z' and a common sequence $y_{0:k} \in \mathcal{Y}$, giving raise to sequences $u_{(-q+1):k+1}$ and $x_{(-p+1):k+1}$ on the one hand, and $u'_{(-q+1):k+1}$ and $x'_{(-p+1):k+1}$ on the other hand. The second line of (2.7) implies that $u'_j = u_j$ for all $j \geq 0$. Thus, using the fourth line and (2.4), we get that, for all $j \geq q$,

$$x_j - x'_j = \sum_{i=1}^p a_i(x_{j-i} - x'_{j-i})$$

Since $\psi^\theta \langle y \rangle(z) = x_{k+1}$ and $\psi^\theta \langle y \rangle(z') = x'_{k+1}$ by (2.6), we thus obtain that, for all $z, z' \in \mathcal{Z}$ and $y \in \mathcal{Y}^{n+1}$, $|\psi^\theta \langle y \rangle(z) - \psi^\theta \langle y \rangle(z')|$ does not depend on y and tends to zero if and only if $a_{1:p} \in \mathcal{S}_p$, where

$$(3.10) \quad \mathcal{S}_p = \left\{ c_{1:r} \in \mathbb{R}^p : \forall z \in \mathbb{C}, |z| \leq 1 \text{ implies } 1 - \sum_{k=1}^p c_k z^k \neq 0 \right\}.$$

We summarize this in the following remark for later reference.

Remark 2. For an LODM(p, q), \mathcal{X} is a closed subset of \mathbb{R} and $U = \mathbb{R}$, with $\delta_U(u, u') = |u - u'|$. Condition (A-3) follows. Moreover, the link function $\psi_y(x)$ is of the form (2.5), which yields that (A-4) is equivalent to the following condition, often referred to in the standard GARCH case as the *invertibility* condition, see [22] for a more general discussion on this topic.

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

Remarkably, all LODMs share the same identifiability condition, which can be expressed as follows, using $a_{1:p}^*$ and $b_{1:q}^*$ to denote the true linear coefficients of the link function (2.5).

(L-2) The polynomials $P_p(\cdot; a_{1:p}^*)$ and $Q_q(\cdot; b_{1:q}^*)$ have no common complex roots, where we defined

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

We have the following result, whose proof can be found in Section 4.5.

Theorem 8. Consider an LODM(p, q) for some $p, q \geq 1$ and let $\theta_* \in \Theta$. Suppose that (A-1) holds with an invariant probability measure satisfying, for all $\theta \in \Theta$,

$$(3.11) \quad \int \ln^+(|\Upsilon(y)|) \pi_V^\theta(dy) < \infty.$$

Suppose moreover that (A-2) and (L-1) hold. Then the following assertions hold for any $\theta_* = (\omega^*, a_{1:p}^*, b_{1:q}^*, \varphi^*) \in \Theta$.

- (i) If (L-2) holds, then the equivalent class $[\theta_*]$ of Definition 4 is reduced to the singleton $\{\theta_*\}$.

(ii) If (L-2) does not hold and there exists a neighborhood of $(\omega^*, a_{1:p}^*, b_{2:q}^*)$ included in $\{(\omega, a_{1:p}, b_{2:q}) : \theta = (\omega, a_{1:p}, b_{1:q}, \varphi) \in \Theta\}$, then $[\theta_\star]$ at least contains a positive length curve going through θ_\star .

Remark 3. If $p = q = 1$, condition (L-2) is reduced to $b_1^* \neq 0$. Let us see what $b_1^* = 0$ would imply about the identifiability of the model in this simple case. Taking ψ^θ as in (2.5) with $p = q = 1$, if $b_1^* = 0$, then $\{X_k : k \in \mathbb{Z}_+\}$ is a deterministic sequence which, under the stationary distribution, has to be constantly equal to $x^* = \frac{\omega^*}{1-a_1^*}$. 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 $b_1 = 0$, $(\omega, a_1) \neq (\omega^*, a_1^*)$ yielding the same constant $\omega/(1-a_1) = \omega^*/(1-a_1^*)$, and such that $G^\theta = G^{\theta^*}$ (because G^θ typically does not depend on (ω, a_1)), we see that the model is not identifiable.

Remark 4. Condition (L-2) holds for “many” parameters $a_{1:p}^*, b_{1:q}^*$, e.g. for Lebesgue almost all ones in \mathbb{R}^{p+q} .

Remark 5. The identifiability condition (L-2) is a well known sufficient condition in the standard GARCH(p, q) models, see [11, (A4)] or [1, Condition (2.27)]. Our result shows that it is much more general and does not apply only in the case where $G^\theta(x, \cdot)$ is a distribution parameterized by a scale parameter x . We moreover show that it is also a necessary condition.

4. POSTPONED PROOFS

4.1. **Proof of Proposition 5.** First observe that (3.1) implies for all $\theta \in \Theta$,

$$(4.1) \quad \tilde{\mathbb{P}}^\theta [Y_1 \in \cdot | Y_{-\infty:0}] = G^\theta(\psi^\theta \langle Y_{-\infty:0} \rangle; \cdot) \quad \tilde{\mathbb{P}}^\theta\text{-a.s.}$$

Let us now show that any $\theta = (\vartheta, \varphi) \in [\theta_\star]$ belongs to $\langle \theta_\star \rangle$, that is, $\varphi = \varphi_\star$ and (3.2) and (3.3) hold true. Since $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta_\star}$, (4.1), which also holds with θ replaced by θ_\star , yields

$$G^\theta(\psi^\theta \langle Y_{-\infty:0} \rangle; \cdot) = G^{\theta_\star}(\psi^{\theta_\star} \langle Y_{-\infty:0} \rangle; \cdot) \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

By (A-2), we obtain that $\theta = (\vartheta, \varphi_\star)$ and (3.2) holds. By Lemma 6, (3.1) implies (3.4), and using $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta_\star}$, we obtain (3.3). Thus $\theta \in \langle \theta_\star \rangle$.

It remains to show that $\langle \theta_\star \rangle \subseteq [\theta_\star]$. Let $\theta \in \langle \theta_\star \rangle$, that is, let $\theta = (\vartheta, \varphi_\star) \in \Theta$ such that (3.2) and (3.3) hold true. Since (3.1) holds with θ replaced by θ_\star , (3.2) gives that $X_1 = \psi^\theta \langle Y_{-\infty:0} \rangle$ $\mathbb{P}^{\theta_\star}$ -a.s. Since $\mathbb{P}^{\theta_\star}$ is shift invariant, we get, for all $k \in \mathbb{Z}$, $X_{k+1} = \psi^\theta \langle Y_{-\infty:k} \rangle$ $\mathbb{P}^{\theta_\star}$ -a.s. With (3.3), we obtain $X_1 = \psi_{Y_{(-q+1):0}}^\theta (X_{(-p+1):0})$ $\mathbb{P}^{\theta_\star}$ -a.s. Since $\mathbb{P}^{\theta_\star}$ is shift invariant, we thus have, for all $k \in \mathbb{Z}$,

$$(4.2) \quad X_{k+1} = \psi_{Y_{(k+1-q):k}}^\theta (X_{(k+1-p):k}) \quad \mathbb{P}^{\theta_\star}\text{-a.s.}$$

On the other hand, by definition of $\mathbb{P}^{\theta_\star}$ and using (A-2) with $\theta = (\vartheta, \varphi_\star)$, we have that

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

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

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

This, with (4.2), shows that $\mathbb{P}^{\theta_\star}$ is a shift-invariant solution of (1.1). By (A-1), we conclude that $\mathbb{P}^{\theta_\star} = \mathbb{P}^\theta$, and thus $\theta \in [\theta_\star]$.

4.2. **Proof of Lemma 7.** We start by introducing some useful notation related to the iterated link function described in Section 2.2. For all $y \in \mathsf{Y}$, we define $\Psi_y^\theta : \mathsf{Z} \rightarrow \mathsf{Z}$ by

$$(4.3) \quad \Psi_y^\theta : z = z_{1:(p+q-1)} \mapsto \left(z_{2:p}, \tilde{\psi}_{u_{1:p}}^\theta(z_{1:p}), z_{(p+2):(p+q-1)}, u_p \right),$$

where $u_{1:(p-1)} = z_{(p+1):(p+q-1)}$ and $u_p = \Upsilon(y)$, and $\tilde{\psi}^\theta$ is the reduced link function appearing in (2.1). The definition (4.3) is only valid for $q > 1$. If $q = 1$, then $\mathsf{Z} = \mathsf{X}^p$ and the definition of Ψ^θ boils down to

$$(4.4) \quad \Psi_y^\theta : z = z_{1:p} \mapsto \left(z_{2:p}, \tilde{\psi}_{u_{1:p}}^\theta(z_{1:p}) \right).$$

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

$$(4.5) \quad \Psi^\theta \langle y_{0:k} \rangle = \Psi_{y_k}^\theta \circ \Psi_{y_{k-1}}^\theta \circ \dots \circ \Psi_{y_0}^\theta.$$

Note that $\psi^\theta \langle y_{0:k} \rangle$ introduced Section 2.2 satisfies

$$(4.6) \quad \psi^\theta \langle y_{0:k} \rangle = \Pi_p \circ \Psi^\theta \langle y_{0:k} \rangle : \mathsf{Z} \rightarrow \mathsf{X}.$$

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

$$(4.7) \quad \Psi^\theta \langle y_{0:k} \rangle(z) = \left((\psi^\theta \langle y_{0:j} \rangle(z))_{k-p < j \leq k}, u_{(k-q+2):k} \right),$$

where we set $u_j = \Pi_{p+q+j}(z)$ for $-q < j \leq -1$ and $u_j = \Upsilon(y_j)$ for $0 \leq j \leq k$ and use the convention $\psi^\theta \langle y_{0:j} \rangle(z) = \Pi_{p-j}(z)$ for $-p < j \leq 0$.

We can now derive the following result.

Lemma 9. (A-4) implies that for all $\theta \in \Theta$, there exists $C > 0$ and $\rho \in (0, 1)$ such that $\text{Lip}_n^\theta \leq C \rho^n$ for all $n \in \mathbb{Z}_+^*$.

Proof. By (3.6), (3.7) and (4.7), we have, for all $n \in \mathbb{Z}_+^*$, using the convention $\text{Lip}_m^\theta = 1$ for $m \leq 0$,

$$(4.8) \quad \sup_{y \in \mathsf{Y}^n, v \in \mathsf{Z}^2} \frac{\delta_{\mathsf{Z}} \circ \Psi^\theta \langle y \rangle^{\otimes 2}(v)}{\delta_{\mathsf{Z}}(v)} \leq \mathbb{1}_{\{n < q\}} \vee \left(\max_{0 \leq j < p} \text{Lip}_{n-j}^\theta \right).$$

Hence (A-4) implies that there exists $m \geq 1$ and $L \in (0, 1)$ such that, for all $y \in \mathsf{Y}^{m+1}$, $\Psi^\theta \langle y \rangle$ is L -Lipschitz. Now observe that, by (4.6), for all $n = km + r$ with $k \geq 0$ and $0 \leq r < m$, for all $y = y_{-n:0} \in \mathsf{Y}^{n+1}$, we can write $\Psi^\theta \langle y \rangle$ as

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

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

$$\delta_{\mathsf{X}}(\psi^\theta \langle y \rangle(z), \psi^\theta \langle y \rangle(z')) \leq \delta_{\mathsf{Z}}(\Psi^\theta \langle y \rangle(z), \Psi^\theta \langle y \rangle(z')) \leq L' L^k \delta_{\mathsf{Z}}(z, z').$$

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

We can now prove Lemma 7. Denote, for all $n \in \mathbb{Z}_+$, $X^{(n)} = \psi^\theta \langle Y_{-n:0} \rangle(z^{(i)})$. Then, by (4.5) we have, for all $n \in \mathbb{Z}_+^*$,

$$X^{(n)} = \psi^\theta \langle Y_{-n+1:0} \rangle \circ \Psi_{Y_{-n}}^\theta(z^{(i)}),$$

and, by (3.6), we get

$$\delta_{\mathsf{X}} \left(X^{(n)}, X^{(n-1)} \right) \leq \text{Lip}_n^\theta \delta_{\mathsf{Z}} \left(z^{(i)}, \Psi_{Y_{-n}}^\theta(z^{(i)}) \right).$$

Using (3.7) with $z^{(i)} = (x^{(i)}, u^{(i)})$ and $x_1^{(i)} = \dots = x_p^{(i)}$ and $u_1^{(i)} = \dots = u_{q-1}^{(i)} = \Upsilon(y_1^{(i)})$ and setting $y^{(i)} = (y_1^{(i)}, \dots, y_1^{(i)}) \in \mathcal{Y}^{q-1}$ we get that

$$\delta_Z \left(z^{(i)}, \Psi_{Y_{-n}}^\theta(z^{(i)}) \right) = \delta_X \left(x_1^{(i)}, \psi_{(y^{(i)}, Y_{-n})}^\theta(x^{(i)}) \right) \vee \delta_U(\Upsilon(y_1^{(i)}), \Upsilon(Y_{-n}))$$

Hence, for all $\alpha > 0$, Condition (3.8) implies, for all $\theta, \theta_* \in \Theta$, as $n \rightarrow \infty$,

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

We conclude with Lemma 9 that (A-4) implies that $\{X^{(n)} : n \in \mathbb{N}\}$ is a Cauchy sequence, hence converges in \mathbf{X} and $\psi^\theta \langle Y_{-\infty:0} \rangle$ is well defined in (3.9) $\tilde{\mathbb{P}}^{\theta_*}, \tilde{\mathbb{P}}^\theta$ -a.s. Moreover, since, as a solution to (1.1) we also have, under \mathbb{P}^θ , for all $n \in \mathbb{Z}_+$, $X_1 = \psi^\theta \langle Y_{-n:0} \rangle(Z_{-n})$, (3.6) also implies

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

Note that $\delta_Z(Z_{-n}, z^{(i)})$ is bounded in probability under \mathbb{P}^θ , hence $X^{(n)}$ converges to X_1 in probability if (A-4) holds. We thus obtain (3.1).

Finally, we check that (3.3) holds when ψ_y^θ is continuous for all $y \in \mathcal{Y}^q$. Since (3.9) holds $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. and $\tilde{\mathbb{P}}^{\theta_*}$ is shift invariant, we have, for all $k \in \mathbb{Z}$,

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

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

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

By continuity of ψ_y^θ and using the previous display, we can take the limit as $n \rightarrow \infty$ under $\tilde{\mathbb{P}}^{\theta_*}$ and obtain (3.3).

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

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

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

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

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

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

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

We further denote by \mathbf{A}_* , \mathbf{b}_* and $\boldsymbol{\omega}_*$ the corresponding values of \mathbf{A} , \mathbf{b} and $\boldsymbol{\omega}$ at $(\omega, a_{1:p}, b_{1:q}) = (\omega^*, a_{1:p}^*, b_{1:q}^*)$, respectively. Recall the notation convention $\theta = (\vartheta, \varphi)$ with $\vartheta = (\omega, a_{1:p}, b_{1:q})$ in [Definition 3](#). We use the corresponding one for θ_* : $\theta_* = (\vartheta_*, \varphi_*)$ with $\vartheta_* = (\omega^*, a_{1:p}^*, b_{1:q}^*)$.

4.4. Useful lemmas. Throughout this section, we use the notation introduced in [Section 4.3](#), in particular for matrix and vectors \mathbf{A} , \mathbf{b} , $\boldsymbol{\varepsilon}_k \dots$. This section contains two lemmas used in the proof of [Theorem 8](#).

Lemma 10. *Let $a_{1:p}^*$ and $b_{1:q}^*$ be in \mathbb{R}^p and \mathbb{R}^q . Suppose that $a_{1:p}^* \in \mathcal{S}_p$, where \mathcal{S}_p is defined in [\(3.10\)](#). Then the following assertions hold.*

(i) *Condition [\(L-2\)](#) implies that, for all $a_{1:p} \in \mathcal{S}_p$ and $b_{2:q} \in \mathbb{R}^{q-1}$,*

$$(4.9) \quad \boldsymbol{\varepsilon}_p^T (\mathbf{A}^k - \mathbf{A}_*^k) \mathbf{b}_* = 0 \quad \text{for all } k \in \mathbb{Z}_+^* \implies \mathbf{A} = \mathbf{A}_*,$$

where $\boldsymbol{\varepsilon}_p$, \mathbf{A} , \mathbf{A}_* and \mathbf{b}_* are defined as in [Section 4.3](#).

(ii) *If, on the contrary, [\(L-2\)](#) does not hold, then there exists $c_{1:p+q-1}^* \in \mathbb{R}^{p+q-1} \setminus \{0\}$ such that for all $a_{1:p} \in \mathcal{S}_p \cap (a_{1:p}^* + \text{Span}(c_{1:p}^*))$ and $b_{2:q} \in b_{2:q}^* + \text{Span}(c_{(p+1):(p+q-1)}^*)$ we have*

$$\boldsymbol{\varepsilon}_p^T (\mathbf{A}^k - \mathbf{A}_*^k) \mathbf{b}_* = 0 \quad \text{for all } k \in \mathbb{Z}_+^*$$

Proof. Let $y = \{y_n : n \in \mathbb{Z}\}$ be a real valued sequence and consider the recursive equation

$$(4.10) \quad x_t = \sum_{k=1}^p a_k x_{t-k} + \sum_{k=1}^q b_k y_{t+1-k}, \quad t \in \mathbb{Z}.$$

If $a_{1:p} \in \mathcal{S}_p$ and y is ℓ^1 (absolutely summable), then there is a unique ℓ^1 solution $x = \{x_n : n \in \mathbb{Z}\}$, given by

$$\hat{x}(\lambda) \left(1 - \sum_{k=1}^p a_k e^{-i\lambda k} \right) = \left(\sum_{k=0}^{q-1} b_{k+1} e^{-i\lambda k} \right) \hat{y}(\lambda),$$

where $\hat{x}(\lambda) = \sum_{k \in \mathbb{Z}} x_k e^{-i\lambda k}$ is the Fourier series of x and \hat{y} is that of y . In particular, if y is the impulse sequence (that is $\hat{y}(\lambda) \equiv 1$), x has a Fourier series given by (using the same notation as in [\(L-2\)](#))

$$(4.11) \quad \hat{x}(\lambda) = \frac{\sum_{k=0}^{q-1} b_{k+1} e^{-i\lambda k}}{1 - \sum_{k=1}^p a_k e^{-i\lambda k}} = e^{i\lambda(p-q-1)} \frac{Q_p(e^{i\lambda}; b_{1:q})}{P_p(e^{i\lambda}; a_{1:p})},$$

Now, writing $\mathbf{z}_t = (x_{(t-p+1):t}, y_{(t-q+2):p})$, Equation [\(4.10\)](#) is equivalent to

$$\mathbf{z}_t = \mathbf{A} \mathbf{z}_{t-1} + y_t \mathbf{b};, \quad t \in \mathbb{Z}.$$

For all $a_{1:p} \in \mathcal{S}_p$, we have $|\lambda|_{\max}(\mathbf{A}) < 1$, and taking y to be the impulse sequence, we have that the unique ℓ^1 solution x must be the causal sequence defined by

$$x_t = \boldsymbol{\varepsilon}_p^T \mathbf{A}^t \mathbf{b}, \quad t \in \mathbb{Z}_+.$$

Hence the left-hand side of (4.9) means that the two equations defined as in (4.10) with coefficients $(a_{1:p}^*, b_{1:q}^*)$ and $(a_{1:p}, b_{1:q}^*)$ have the same ℓ^1 solution when y is impulse sequence, hence by (4.11), it is equivalent to the polynomial equation

$$Q^* \times P = P^* \times Q ,$$

where we introduced the polynomials $P^*(z) = P_p(z; a_1^*, \dots, a_p^*)$, $Q^*(z) = Q_q(z; b_1^*, \dots, b_q^*)$, $P(z) = P_p(z; a_1, \dots, a_p)$ and $Q(z) = Q_q(z; b_1^*, b_2, \dots, b_q)$.

We deduce from this analysis and using these definitions that, to prove Lemma 10 (i), we only need to show that, whenever $a_{1:p}^*, a_{1:p} \in \mathcal{S}_p$, if Condition (L-2) holds, then

$$Q^*P = P^*Q \text{ implies } Q = Q^* \text{ and } P = P^* .$$

This is indeed true, since, under (L-2), by the Gauss Theorem, if $Q^*P = P^*Q$ implies that P^* divides P and, since they have the same degree p and are unitary, we get $P = P^*$, from which it follows that $Q = Q^*$.

Next, we prove Lemma 10 (ii). Using the analysis above, we suppose that P^* and Q^* have at least one common root and show that we can find two real polynomials C and D , with C non-zero (and $D = 0$ only in the special case $q = 1$), with degrees at most $p - 1$ and $q - 2$ such that, for all $d \in \mathbb{R}$,

$$(4.12) \quad Q^*(P^* + d C) = P^*(Q^* + d D) .$$

The constraints on the degrees of C and D guaranty that $P^* + d C$ and $Q^* + d D$ can be expressed as P and Q above for some $a_{1:p} \in a_{1:p}^* + \text{Span}(c_{1:p}^*)$ and $b_{2:q} \in b_{2:q}^* + \text{Span}(c_{(p+1):(p+q-1)}^*)$, where $c_{1:p}^*$ and $c_{(p+1):(p+q-1)}^*$ are coefficients in the polynomials C and D . Let U denote the greater common divisor of P^* and Q^* , which has degree at least one by assumption. Now define C and D by $Q^* = D \times U$ and $P^* = C \times U$. Then the degrees of C and D are at most $p - 1$ and $q - 2$, as requested, and (4.12) can be readily checked. \square

Lemma 11. *Define ε_p and \mathbf{A} as in Section 4.3 and suppose that $a_{1:p} \in \mathcal{S}_p$, where \mathcal{S}_p is defined in (3.10). Then $\sum_{k \geq 0} \varepsilon_p^T \mathbf{A}^k \varepsilon_p \neq 0$.*

Proof. We have, for all $k \geq 0$, $\varepsilon_p^T \mathbf{A}^k \varepsilon_p = \tilde{\varepsilon}_p^T \tilde{\mathbf{A}}^k \tilde{\varepsilon}_p$, where $\tilde{\varepsilon}_p$ denotes the last canonical vector of \mathbb{R}^p and $\tilde{\mathbf{A}}$ is the companion matrix defined as the $p \times p$ upper left bloc of \mathbf{A} . Since $a_{1:p} \in \mathcal{S}_p$, we have $\sum_{k \geq 0} \tilde{\mathbf{A}}^k = (\mathbf{I}_p - \tilde{\mathbf{A}})^{-1}$. To obtain the result we observe that $\tilde{\varepsilon}_p^T (\mathbf{I}_p - \tilde{\mathbf{A}})^{-1} \tilde{\varepsilon}_p = 0$ implies that there exists $c_1, \dots, c_{p-1} \in \mathbb{R}$ such that $\tilde{\varepsilon}_p = (\mathbf{I}_p - \tilde{\mathbf{A}}) [c_1 \ \dots \ c_{p-1} \ 0]^T$. Using the companion matrix form of $\tilde{\mathbf{A}}$, and looking successively at the $p - 1, p - 2, \dots$ entries in the previous equation, we easily get iteratively that $c_{p-1} = 0, \dots, c_1 = 0$, and thus obtain a contradiction. \square

We can now prove the main theorem.

4.5. Proof of Theorem 8. By Remark 2, (A-4) holds as a consequence of (L-1). We apply Proposition 5. To this end we must first show that (3.1) holds. For this we apply Lemma 7. Take an arbitrary $x_1^{(i)} \in \mathcal{X}$, and if $q > 1$, $y^{(i)} \in \mathcal{Y}^{q-1}$. Then, since $\psi_y^\theta(x)$ is of the form (2.5), there exists constants $C_1, C_2 > 0$ only depending on $\theta, x_1^{(i)}$ and $y^{(i)}$ such that, for all $y \in \mathcal{Y}$,

$$\delta_{\mathcal{X}} \left(x_1^{(i)}, \psi_{(y^{(i)}, y)}^\theta(x^{(i)}) \right) \leq C_1 + C_2 |\Upsilon(y)| .$$

Thus (3.11) implies that (3.8) holds and we can apply Lemma 7 to get (3.1) and (3.9). Using the notation of Section 4.3, we have for all $k \geq 0$,

$$Z_k = \boldsymbol{\omega} + \mathbf{A}Z_{k-1} + \mathbf{b}U_{k-1}$$

where Z_k is defined in (2.2). Then, noting that $X_k = \boldsymbol{\varepsilon}_p^T Z_k$, (3.9) becomes

$$(4.13) \quad \psi^\theta \langle Y_{-\infty:0} \rangle = \sum_{k=0}^{\infty} \boldsymbol{\varepsilon}_p^T \mathbf{A}^k (\boldsymbol{\omega} + \Upsilon(Y_{-k}) \mathbf{b}) \quad \tilde{\mathbb{P}}^{\theta_*}, \tilde{\mathbb{P}}^\theta\text{-a.s.}$$

Now we can prove Assertion (i) of Theorem 8. Suppose that (L-2) holds and take $\theta \in [\theta_*]$. By Proposition 5, this implies $\theta \in \langle \theta_* \rangle$ and thus $\theta = (\vartheta, \varphi_*)$ and satisfies (3.2). It remains to show that $\vartheta = \vartheta_*$. Since (4.13) holds both $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. and $\tilde{\mathbb{P}}^\theta$ -a.s., and the same equation holds with θ , \mathbf{A} , $\boldsymbol{\omega}$ and \mathbf{b} replaced by θ_* , \mathbf{A}_* , $\boldsymbol{\omega}_*$ and \mathbf{b}_* , (3.2) implies

$$(4.14) \quad \sum_{k=0}^{\infty} \boldsymbol{\varepsilon}_p^T \mathbf{A}^k (\boldsymbol{\omega} + \Upsilon(Y_{-k}) \mathbf{b}) = \sum_{k=0}^{\infty} \boldsymbol{\varepsilon}_p^T \mathbf{A}_*^k (\boldsymbol{\omega}_* + \Upsilon(Y_{-k}) \mathbf{b}_*) \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

This implies that $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.,

$$(b_1^* - b_1) \Upsilon(Y_0) = \boldsymbol{\omega} - \boldsymbol{\omega}_* + \sum_{k=1}^{\infty} \boldsymbol{\varepsilon}_p^T \mathbf{A}^k (\boldsymbol{\omega} + \Upsilon(Y_{-k}) \mathbf{b}) - \sum_{k=1}^{\infty} \boldsymbol{\varepsilon}_p^T \mathbf{A}_*^k (\boldsymbol{\omega}_* + \Upsilon(Y_{-k}) \mathbf{b}_*)$$

If $b_1 \neq b_1^*$, this implies that $\Upsilon(Y_0)$ is a deterministic function of $Y_{-\infty:1}$. Under \mathbb{P}^{θ_*} , by (3.1), the distribution of Y_0 given $Y_{-\infty:1}$ is $G^{\theta_*}(X_1, \cdot)$. By (1.2) $G^{\theta_*}(X_1, \cdot)$ has the same support as ν . Hence the distribution of Y_0 given $Y_{-\infty:1}$ has the same support as $\nu \circ \Upsilon^{-1}$ which is not degenerate by Definition 3-(ii), and we get a contradiction. So we must have $b_1 = b_1^*$, that is, $\mathbf{b} = \mathbf{b}_*$. Iterating, and using this line of reasoning, we get that, for all $k \geq 1$,

$$\boldsymbol{\varepsilon}_p^T \mathbf{A}^k \mathbf{b}_* = \boldsymbol{\varepsilon}_p^T \mathbf{A}_*^k \mathbf{b}_* .$$

Applying Lemma 10, we have that this implies $\mathbf{A} = \mathbf{A}_*$, that is, $a_{1:p} = a_{1:p}^*$ and $b_{2:q} = b_{2:q}^*$. Going back to (4.14), we also get that

$$\sum_{k=0}^{\infty} \boldsymbol{\varepsilon}_p^T \mathbf{A}_*^k \boldsymbol{\varepsilon}_p (\boldsymbol{\omega} - \boldsymbol{\omega}_*) = 0 ,$$

and we conclude with Lemma 11 that $\boldsymbol{\omega} = \boldsymbol{\omega}_*$. Collecting the obtained identities, we arrive at $\vartheta = \vartheta_*$ and this concludes the proof of Theorem 8 (i).

Next, we prove Theorem 8 (ii). Hence we suppose that (L-2) does not hold, and, since we want to show that $[\theta_*]$ contains a particular set, we use that $[\theta_*] = \langle \theta_* \rangle$ as given by Proposition 5. Hence, we only need to exhibit $\theta = (\vartheta, \varphi_*)$ such that (3.2) and (3.3) hold. By Lemma 7, (3.3) holds for all $\theta \in \Theta$ and we only need to impose (3.2) on ϑ . Using (4.13), which holds $\tilde{\mathbb{P}}^\theta, \tilde{\mathbb{P}}^{\theta_*}$ -a.s. for all θ , including $\theta = \theta_*$, it is thus sufficient to have

$$\text{for all } k \geq 0, \boldsymbol{\varepsilon}_p^T \mathbf{A}^k \mathbf{b} = \boldsymbol{\varepsilon}_p^T \mathbf{A}_*^k \mathbf{b}_* \quad \text{and} \quad \sum_{k=0}^{\infty} \boldsymbol{\varepsilon}_p^T \mathbf{A}^k \boldsymbol{\omega} = \sum_{k=0}^{\infty} \boldsymbol{\varepsilon}_p^T \mathbf{A}_*^k \boldsymbol{\omega}_* .$$

The first relation with $k = 0$ corresponds to $b_1 = b_1^*$, which we assume in the following, so that $\mathbf{b} = \mathbf{b}_*$. By Lemma 10, we can find a line of \mathbb{R}^{p+q-1} , going through $(a_{1:p}^*, b_{2:q}^*)$ such that these relations hold for all $k \geq 0$, for all $(a_{1:p}, b_{2:q})$ on

this line. Now, by definition of ω_* and ω , and using [Lemma 11](#) with $a_{1:p} \in \mathcal{S}_p$, the last relation in the displayed sufficient condition can be written as

$$\omega = \frac{\sum_{k=0}^{\infty} \varepsilon_p^T \mathbf{A}_*^k \varepsilon_p}{\sum_{k=0}^{\infty} \varepsilon_p^T \mathbf{A}^k \varepsilon_p} \omega^*,$$

where the ratio tends to 1 as $(a_{1:p}, b_{2:q})$ approaches $(a_{1:p}^*, b_{2:q}^*)$. Hence we get that $[\theta_*]$ contains the part of the curve obtained by taking $(a_{1:p}, b_{2:q})$ on the previously exhibited line and sufficiently close to $(a_{1:p}^*, b_{2:q}^*)$ to have $a_{1:p} \in \mathcal{S}_p$, and setting ω as above.

REFERENCES

- [1] István Berkes, Lajos Horváth, and Piotr Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, 9(2):201–227, 2003. ISSN 1350-7265. doi: 10.3150/bj/1068128975. URL <https://doi.org/10.3150/bj/1068128975>.
- [2] 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.
- [3] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31:307–327, 1986.
- [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.
- [6] DR Cox. Statistical analysis of time-series: some recent developments. *Scand. J. Statist.*, 8(2):93–115, 1981. ISSN 0303-6898.
- [7] Richard A Davis, WTM Dunsmuir, and Y Wang. Modeling time series of count data. *STATISTICS TEXTBOOKS AND MONOGRAPHS*, 158:63–114, 1999.
- [8] 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.
- [9] 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.
- [10] K. Fokianos and D. Tjøstheim. Log-linear poisson autoregression. *J. of Multivariate Analysis*, 102(3):563–578, 2011.
- [11] Christian Francq and Jean-Michel Zakoian. Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4):605–637, 2004.
- [12] Christian Francq and Jean-Michel Zakoian. A tour in the asymptotic theory of garch estimation. *Handbook of Financial Time Series*, pages 85–111, 2009.
- [13] Christian Francq and Jean-Michel Zakoian. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2011.
- [14] 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.
- [15] B. G. Leroux. Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.*, 40:127–143, 1992.

- [16] Roman Liesenfeld and Jean-Francois Richard. Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of empirical finance*, 10(4):505–531, 2003.
- [17] 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.
- [18] Robert S Pindyck and Daniel L Rubinfeld. *Econometric models and economic forecasts*, volume 4. Irwin/McGraw-Hill Boston, 1998.
- [19] 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.
- [20] 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>.
- [21] 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>.
- [22] D. Straumann and T. Mikosch. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *ArXiv Mathematics e-prints*, February 2007.
- [23] Dag Tjøstheim. *Count Time Series with Observation-Driven Autoregressive Parameter Dynamics*. Handbook of Discrete-Valued Time Series. London: Chapman & Hall, 2015.
- [24] Scott L Zeger. A regression model for time series of counts. *Biometrika*, 75(4):621–629, 1988.
- [25] 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>.

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

LTCI, TÉLÉCOM PARISTECH, UNIVERSITÉ PARIS-SACLAY, 75013 PARIS, FRANCE
E-mail address: roueff@telecom-paristech.fr

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