Stationarity and ergodicity of Markov switching positive conditional mean models

Aknouche, Abdelhakim and Francq, Christian

USTHB and Qassim University, CREST and University of Lille

18 August 2020

Stationarity and ergodicity of Markov switching positive conditional mean models

Abdelhakim Aknouche

USTHB and Qassim University

and

Christian Francq*

CREST and University of Lille

Abstract

A general Markov-Switching *autoregressive conditional mean* model, valued in the set of nonnegative numbers, is considered. The conditional distribution of this model is a finite mixture of nonnegative distributions whose conditional mean follows a GARCH-like dynamics with parameters depending on the state of a Markov chain. Three different variants of the model are examined depending on how the lagged-values of the mixing variable are integrated into the conditional mean equation. The model includes, in particular, Markov mixture versions of various well-known nonnegative time series models such as the autoregressive conditional duration (ACD) model, the integer-valued GARCH (INGARCH) model, and the Beta observation driven model. Under contraction in mean conditions, it is shown that the three variants of the model are stationary and ergodic when the stochastic order and the mean order of the mixing distributions are equal. The proposed conditions match those already known for

*I am grateful to the Agence Nationale de la Recherche (ANR), which supported this work via the Project MultiRisk (ANR-16-CE26-0015-02). I also thanks the labex ECODEC.

Markov-switching GARCH models. We also give conditions for finite marginal moments. Applications to various mixture and Markov mixture count, duration and proportion models are provided.

Keywords: Autoregressive Conditional Duration, Count time series models, finite mixture models, Ergodicity, Integer-valued $GARCH$, Markov mixture models.

1 Introduction

Nonnegative time series generally refer to: i) $(0, \infty)$-valued data such as durations and volumes, ii) integer-valued series, namely count and binary data, and iii) bounded-valued observations including proportions, indices, probabilities and rates. Numerous models have been introduced in recent years to model nonnegative time series, according to the Fisher principle that data are better described in their "natural habitat" (cf. Fisher, 1953; Jorgensen, 1997). They specifically aim to reproduce data in their actual type, without transforming them to be modeled by linear or Gaussian ARMA-like models. Adopting the Generalized Linear Model (GLM, Nelder and Wedderburn, 1972) approach which proposes a unified framework to deal with a vast class of non-Gaussian data using exponential distribution family (cf. McCullag and Nelder, 1989), recent time series models aim to go beyond the independent-data assumption (Benjamin et *al*, 2003) or/and the exponential family framework (Zheng et *al*, 2015).

For positive real $(0, \infty)$-valued time series, one of the first and best known time series model is the autoregressive conditional duration (ACD) introduced by Engle and Russell (1998). Although designed to rather model the conditional mean, the ACD model, in its original form, has a similar multiplicative error model (MEM, Engle, 2002) structure as the GARCH (generalized autoregressive conditional heteroskedastic) equation, and therefore its ergodic properties are obtained in the same way through the stochastic recurrence equation theory (SRE, Bougerol, 1993). The MEM form of the ACD is, however, quite restrictive since the conditional variance must be proportional to the squared conditional mean and the innovation term must be independent of past observations (Aknouche and Francq, 2020). In particular, MEM cannot accommodate certain interesting distributions that do not belong to the exponential family, such as the Gamma distribution with time-varying shape parameter (Zheng et *al*, 2015; Bhogal and Variyam, 2019). To avoid the constraints imposed by the MEM form, extensions of the ACD model have been defined through the specification of a positive-valued conditional distribution whose mean depends on past values (Creal et *al*, 2013; Harvey, 2013; Zheng et *al*, 2015; Aknouche and Francq, 2020). The study of the ergodic structure of such an extended ACD via the SRE theory is however not obvious because the model is not explicitly defined by means of an iid (independent and identically distributed) white noise. The same issue arises for the integer-valued GARCH model (INGARCH, Rydberg and Shephard, 2000; Heinen, 2003; Ferland et *al*, 2006) which has the same structure as the extended ACD model, but where the conditional distribution is discrete. A large amount of effort has been executed recently to study the ergodic properties of INGARCH models and their various extensions (e.g. Fokianos et *al*, 2009; Gonçalves et *al*, 2015; Davis and Liu, 2016). Aknouche and Francq (2020) proposed ergodicity conditions for a general positive conditional mean model which includes in particular the ACD and INGARCH models and whose conditional distribution belongs to a vast class of distributions for which the (conditional) stochastic and mean orders are the same. Such a class encompasses the one-parameter exponential family and other interesting distributions with time varying shape, scale or rate parameters. Recently, Gorgi and Koopman (2020) adopted a similar approach to establish the ergodicity of a Beta observation-driven process which is conditionally Beta distributed with a time-varying mean having a similar ACD (and thus an INGARCH) equation.

Since the pioneering work by Hamilton (1989), the Markov Switching (MS) formulation has been integrated into many real-valued time series models (e.g. ARMA, GARCH and bilinear time series models) in order to take into account certain specific observed facts, namely recurrent changes in regime, multimodality, and heavy-tailness of the marginal distribution (see e.g. Hamilton and Susmel, 1994; Bauwens et *al*, 2010; Francq and Zakoian, 2019 and the references therein). The MS device for a given equation consists in considering time-varying parameters depending on the state of a finite stationary and ergodic Markov chain. In the context of positive data, various Markov switching ACD (MS-ACD) models have been introduced to model realized volatility in microstructure markets (e.g. Hujer et *al*, 2002, 2003; De Luca and Zuccolotto, 2006; Hujer and Vuletic, 2007; Hautsch, 2012; Chen et *al*, 2013). Recently, Markov-switching INGARCH (MS-INGARCH) models have also been introduced in the context of count time series (Zhu et *al*, 2010; Berentsen et *al*, 2018; Lee and Hwang, 2018; Doukhan et *al*, 2018; Aknouche and Demmouche, 2019). The Markov switching count and duration models have a similar conditional mean specification as MS-GARCH models, but unlike the latter, they are generally not MEM and therefore their ergodic properties are not easily obtained using the SRE theory. In spite of the various existing MS-ACD formulations it seems that no stationarity and ergodicity results have been established. For count time series, stationarity and ergodicity of MS-INGARCH models with independent switching has been studied by Aknouche and Demmouche (2019), Doukhan et *al* (2018), and Mao et *al* (2019). These results were based on the weak dependence approach of Doukhan and Wintenberger (2008) which only works for iid switching. In addition, the ergodicity conditions proposed are generally not necessary and may therefore be refined.

In this paper we propose stationarity, ergodicity and finite moment conditions for a unified class of Markov switching positive conditional mean models in which the coefficients of the conditional mean equation are allowed to depend on the state of a finite unobserved stationary and ergodic Markov chain. The corresponding mixing distributions are assumed to belong to the class of equal conditional stochastic and mean orders introduced by Aknouche and Francq (2020). Three different formulations are considered, differing in the way the past and present of the regime variable is integrated into the conditional mean dynamics. The first one we call *past-regime dependent switching* considers the lagged values of the conditional mean depending on the lagged values of the regime variable (e.g. Aknouche and Demmouche, 2019; Doukhan et *al*, 2018; Mao et *al*, 2019). This resembles the formulation by Francq and Roussignol (1998). In the second one, which is named the *present-regime dependent switching* and is inspired by Fong and See, (2001), Hujer and Vuletic (2002) and

Haas et *al* (2004), the lagged conditional mean values are governed by the present of the regime variable (see Diop et *al*, 2016-2018; Berentsen et *al*, 2018 for MS-INGARCH models). In the third one we call *present-regime mean-dependent switching*, which is similar to the earlier formulation proposed by Gray (1996) for MS-GARCH models, the lagged values of the time-varying parameter depends on the present conditional mean of the model (see also Hujer and Vuletic, 2007 for an MS-ACD model and Lee and Hwang, 2018 for their MS-INGARCH model).

The rest of this paper is organized as follows. Section 2 reviews the general formulation of finite mixture and Markov mixture models, and propose the three versions of the Markov-switching positive conditional mean model. Ergodicity and finite moment conditions are given in Section 3 for the past-regime dependent switching, in Section 4 for the present-regime dependent switching and in Section 5 for the present-regime mean-dependent switching. Applications to existing specific Mixture and Markov count and duration models are provided. Section 6 compares the theoretical means that we obtained with empirical means of Monte Carlo simulations and Section 7 concludes, while the proofs of the main results are left to the Appendix.

2 Markov-switching positive conditional mean models

Let $\mathcal{N} \subset [0, \infty)$ be a subset of nonnegative real numbers which may refer to $[0, \infty)$ itself, to the set of positive real numbers $(0, \infty) = \mathbb{R}_{+}$, to the set of integer numbers $\mathbb{N} = \{0, 1, ...\}$, or to any bounded interval $[a, b]$ of the real line $(0 \leq a < b)$, where a special case is played by the simplex $[0, 1]$ interval. In the sequel, all observable stochastic processes of interest are defined on a probability space $(\Omega, \mathcal{F}, P)$ and valued in $\mathcal{N}$. Let F_{λ} be a cumulative distribution function (cdf) indexed by the mean $\lambda = \int_{0}^{+\infty} x dF_{\lambda}(x) > 0$ with support $\mathcal{N}$. A stochastic process $\{Y_t, t \in \mathbb{Z}\}$ $(\mathbb{Z} = \{..., -1, 0, 1, ...\})$ is said to be a positive linear conditional mean (POLI) model if its conditional distribution is F_{λ_t}, that is

$$Y_t \mid \mathcal{F}_{t-1} \sim F_{\lambda_t} \tag{2.1}$$

with conditional mean satisfying

$$\lambda_t = \omega + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j \lambda_{t-j}, \tag{2.2}$$

where p and q are nonnegative integers, and $\mathcal{F}_t$ denotes the information set available at time t, i.e. the sigma-field generated by $\{Y_u, u \leq t\}$. To ensure the positivity of the conditional mean we take $\omega > 0$, $\alpha_i \geq 0$ and $\beta_j \geq 0$ $(i = 1, ..., p,\ j = 1, ..., q)$. The POLI model (2.1)-(2.2) reduces to the extended ACD model (Aknouche and Francq, 2020) when F_λ has as support $\mathcal{N} = (0, \infty)$ or $[0, \infty)$[1], to the INGARCH(p, q) model (Rydberg and Shephard, 2000; Heinen, 2003) when F_λ has a discrete support $\mathcal{N} = \mathbb{N}$, and to a bounded-valued GARCH model when $\mathcal{N} = [a, b]$. See e.g. Gorgi and Koopman (2020) for the Beta observation driven model with F_λ being the beta distribution and $\mathcal{N} = [0, 1]$. The POLI model also includes positive-valued versions of the generalized autoregressive score (GAS) model of Creal et *al* (2013) and Harvey (2013).

Assume the cdf F_λ belongs to the class $\mathbb{F}$ of distributions with equal conditional stochastic and mean orders (cf. Aknouche and Francq, 2020) which satisfies the following property

$$\lambda \leq \lambda^* \quad \Rightarrow \quad F_\lambda^-(u) \leq F_{\lambda^*}^-(u),\ \forall u \in (0, 1), \tag{2.3}$$

where F_λ^- is the quantile function (generalized inverse) associated to F_λ. This rich class of distributions whose stochastic order is driven by the mean includes in particular the one-parameter exponential family. Important distributions that do not belong to the exponential family, such as the negative binomial distributions with time-varying numbers of failures, the Gamma distributions with time-varying shape parameters, the Beta distribution with time varying parameters, and the zero-inflated Poisson distribution, are also in $\mathbb{F}$.

We now consider three Markov-switching generalizations of the POLI model. Let a positive integer S. The process $\{Y_t, t \in \mathbb{Z}\}$ is said to be a switching (or mixture) POLI model if its conditional distribution given the past information is a mixture of distributions,

$$Y_t \mid \mathcal{F}_{t-1} \sim \pi_{1t} F_{1,\lambda_{1t}} + \cdots + \pi_{St} F_{S,\lambda_{St}}, \tag{2.4}$$

[1]In particular when F_λ satisfies: $F_\lambda(x) = F_1(x/\lambda)$ for all x, the POLI model reduces to the standard ACD (Engel and Russell, 1998).

where $\pi_{st} \geq 0$, $\sum_{s=1}^{S} \pi_{st} = 1$ $(t \in \mathbb{Z})$, and $F_\lambda = F_{s,\lambda}$ belongs to the class of distributions $\mathbb{F}$ given by (2.3). Three dynamics for λ_{st} $(1 \leq s \leq S)$ are to be considered below. The subscript s in (2.4) denotes the regime and S stands for the number of regimes. It is not necessary that all mixing distributions $F_{s,\lambda_{st}}$ be the same across s. For instance, in the mixture INGARCH model, we can for example mix the Poisson and the negative binomial distributions. Let a stationary regime sequence $\{\Delta_t, t \in \mathbb{Z}\}$ be defined on $(\Omega, \mathcal{F}, P)$ and valued in the set $\{1, ..., S\}$ with

$$\pi_s = P\left(\Delta_t = s\right) \quad \text{and} \quad \pi_{st} = P\left(\Delta_t = s \mid \mathcal{F}_{t-1}\right), \quad 1 \leq s \leq S. \tag{2.5}$$

where $\pi_s \geq 0$, $\sum_{s=1}^{S} \pi_s = 1$. Any dependence structure assumed for $\{\Delta_t, t \in \mathbb{Z}\}$ determines a specific type of switching. The sequence $\{\Delta_t, t \in \mathbb{Z}\}$ may be iid, a case we refer to as the iid switching (or iid mixture) for which (2.5) simply writes

$$\pi_s = \pi_{st} \quad 1 \leq s \leq S, \ t \in \mathbb{Z}.$$

It may also be a stationary and ergodic Markov chain with transition probability

$$p_{ij} = P\left(\Delta_t = j \mid \Delta_{t-1} = i\right), \quad i, j \in \{1, ..., S\}.$$

In such a case, we call the model Markov-switching (or Markov mixture) POLI (MS-POLI) where π_{st} can be obtained in terms of (p_{ij}) and (π_s) (e.g. Hamilton, 1994). In the latter cases, the sequence $\{\Delta_t, t \in \mathbb{Z}\}$ is generally assumed unobservable. Another notable but degenerate case is the threshold mixture in which Δ_t is observable and depends on $\mathcal{F}_{t-1}$. Denote by $\mathcal{F}_t^a$ the "complete" (or augmented) information set, the sigma-field generated by $\{Y_u, \Delta_{u+1}, u \leq t\}$. The conditional distribution in (2.4) can be rewritten in term of Δ_t

$$Y_t \mid \mathcal{F}_{t-1}^a \sim F_{\Delta_t, \lambda_{\Delta_t, t}}. \tag{2.6}$$

The specification of a dynamics for $\lambda_{\Delta_t,t}$ in (2.6) is an issue when $\lambda_{\Delta_t,t}$ has to be lagged as it happens for MS models with moving average or GARCH-like components (e.g. Gray, 1996; Fong and See, 2001; Klaassen, 2002; Francq and Zakoian, 2001, 2008; Haas et *al*, 2004).

A natural specification for the conditional mean mixture is

$$\lambda_{\Delta_t,t} = \omega_{\Delta_t} + \sum_{i=1}^{q} \alpha_{\Delta_t,i} Y_{t-i} + \sum_{j=1}^{p} \beta_{\Delta_t,j} \lambda_{\Delta_{t-j},t-j}, \tag{2.7}$$

where $\omega_s > 0, \alpha_{si} \geq 0$ and $\beta_{sj} \geq 0$ $(1 \leq s \leq S)$. A more general specification has the form

$$\lambda_t = g_{\Delta_t}(Y_{t-1}, \ldots, Y_{t-q}, \lambda_{t-1}, \ldots, \lambda_{t-p}), \tag{2.8}$$

where the functions g_s for $1 \leq s \leq S$ are valued in $[0, \infty)$ and $\lambda_t = \lambda_{\Delta_t,t}$.

The lagged values of $\lambda_{\Delta_t,t}$ in (2.7) thus depend on the lagged values in same order of the regime variable Δ_t (see also Francq and Zakoian, 2005 for the MS-GARCH). For MS-INGARCH models with independent switching, representation (2.7) was considered by Aknouche and Demmouche (2019), Doukhan et *al* (2018), and Mao et *al* (2019). The likelihood of (2.6)-(2.7) based on a series $Y_1, ..., Y_n$ is not easy to compute because it depends on the whole path history of Δ_t. To remedy this drawback, one can consider the following specification due to Hujer et *al* (2002) and also introduced by Fong and See (2001) and Haas et *al* (2004) in the context of the real-valued MS-GARCH model

$$\lambda_{st} = \omega_s + \sum_{i=1}^{q} \alpha_{si} Y_{t-i} + \sum_{j=1}^{p} \beta_{sj} \lambda_{s,t-j}, \quad 1 \leq s \leq S \tag{2.9}$$

or more generally

$$\lambda_{st} = g_s(Y_{t-1}, \ldots, Y_{t-q}, \lambda_{s,t-1}, \ldots, \lambda_{s,t-p}), \quad 1 \leq s \leq S \tag{2.10}$$

for some $[0, \infty)$-valued functions g_s, $1 \leq s \leq S$. Differently from (2.7), the specifications (2.9) and (2.10) are such that all lagged values of $\lambda_{\Delta_t,t}$ are governed by the present regime Δ_t, i.e. $\lambda_{\Delta_t,t}$ is a function of $\lambda_{\Delta_t,t-1}, ..., \lambda_{\Delta_t,t-p}$ for all fixed $\Delta_t \in \{1, ..., S\}$. Thus (2.9)-(2.10) are useful for expectation-maximization likelihood purposes (Hamilton, 1989, 1994, Haas et *al*, 2004) but in some respects they seem artificial in describing regime switching. In particular, with this model it seems difficult to guarantee strong persistence of the conditional mean. Suppose for instance that the regime $\Delta_t = 1$ corresponds to a low level of conditional mean and that the regime $\Delta_t = 2$ corresponds to a high level of conditional mean: $0 \simeq \lambda_{1,t} << \lambda_{2,t}$.

If at time $t-1$ the actual conditional mean was high ($\lambda_{t-1} = \lambda_{2,t-1} >> 0$) and if the regime changes at time t, then λ_t cannot remain at a high level, even if the "persistence" parameter β is close to 1, because λ_t depends on the "virtual" past value of the conditional mean $\lambda_{1,t-1} \simeq 0$, and not on its actual past value λ_{t-1}.

A similar representation to (2.9) based on Gray (1996) assumes that

$$\lambda_{st} = \omega_s + \sum_{i=1}^{q} \alpha_{si} Y_{t-i} + \sum_{j=1}^{p} \beta_{sj} E\left(Y_{t-j} \mid \mathcal{F}_{t-j-1}\right), \quad 1 \leq s \leq S, \tag{2.11}$$

where $E\left(Y_t \mid \mathcal{F}_{t-1}\right)$ is the conditional mean of the model. The specification (2.11) has been proposed by Hujer et al (2002, 2003) for their MS-ACD models and adopted by Lee and Hwang (2018) in their MS-INGARCH model. Following Klaassen (2002) in the MS-GARCH case, we replace all cases of $E\left(Y_{t-j} \mid \mathcal{F}_{t-j-1}\right)$ in (2.11) by $E\left(\lambda_{\Delta_{t-j},t-j} \mid \Delta_t = s, \mathcal{F}_{t-1}\right)$, leading to

$$\lambda_{st} = \omega_s + \sum_{i=1}^{q} \alpha_{si} Y_{t-i} + \sum_{j=1}^{p} \beta_{sj} E\left(\lambda_{\Delta_{t-j},t-j} \mid \Delta_t = s, \mathcal{F}_{t-1}\right). \tag{2.12}$$

Klaassen (2002) argued that the analog specification of (2.12) in the MS-GARCH case allows a better multi-step ahead forecasting performance, simplifies the computation implied by the specification (2.11) of Gray (1996) and is more flexible in generating persistence. An extension of (2.12) to nonlinear forms is

$$\lambda_{st} = g_s(Y_{t-1}, \ldots, Y_{t-q}, \mu_{t-1}, \ldots, \mu_{t-p}), \quad 1 \leq s \leq S, \tag{2.13}$$

where

$$\mu_{t-j} = E\left(\lambda_{\Delta_{t-j},t-j} \mid \Delta_t = s, \mathcal{F}_{t-1}\right), \ \ j = 1, ..., p$$

and g_s $(1 \leq s \leq S)$ is defined as above.

We refer to (2.7), (2.9) and (2.12) as, respectively, the past-regime dependent switching, the present-regime dependent switching and the present-regime mean-dependent switching. These various specifications are adaptations of the ones proposed for the Markov-switching GARCH models. Due to the non-multiplicative error form of the general MS-POLI model (2.6) with dynamics (2.7), (2.9) or (2.12), its ergodic structure is more difficult to reveal than MEM-like MS-GARCH models. Sections 3, 4 and 5 aspire to solve this problem.

In the sequel, (Δ_t) denotes an irreducible, stationary and ergodic Markov chain with finite state space $\{1,2,\ldots,S\}$. Let $p_{ij}^{(l)} = P(\Delta_{t+l} = j \mid \Delta_t = i)$ be the l-step transition probabilities, and $p_{ij}^{(1)} = p_{ij}$. The processes (Y_t) and (λ_t) that we will consider are causal functions of (Δ_t) and of another latent process (U_t) that is assumed to be independent of (Δ_t). Under this assumption, denoting by $\mathcal{I}_t$ the sigma-field generated by $\{U_s, \Delta_s : s \leq t\}$, the variable Δ_t contains all the information of $\mathcal{I}_t$ that is useful to predict the future values of the chain, in the sense that

A1 $P(\Delta_t = s \mid \Delta_{t-1} = \tau, A) = P(\Delta_t = s \mid \Delta_{t-1} = \tau)$ for all event $A \in \mathcal{I}_{t-1}$.

3 The past-regime dependent switching

We give stationarity and ergodicity conditions for the past-regime dependent MS-POLI model given by (2.6) and (2.7). We also study the existence of marginal moments in the case $p = q = 1$. Specific conditions when (Δ_t) is iid are obtained and illustrations of the general results on MS-ACD and MS-INGARCH models are provided. We then extend the stationarity results to the nonlinear conditional mean form (2.8).

3.1 The linear conditional mean case

Define the $S \times S$ matrix $M^{(l)}$ for $l = 1, ..., r = \max(p,q)$ by

$$M^{(l)}(i,j) = p_{ji}^{(l)}(\alpha_{il} + \beta_{il}), \quad i,j \in \{1, ...S\}$$

with obvious notations. For instance,

$$M^{(1)} = \begin{pmatrix} p_{11}(\alpha_{11}+\beta_{11}) & p_{21}(\alpha_{11}+\beta_{11}) & \cdots & p_{S1}(\alpha_{11}+\beta_{11}) \\ p_{12}(\alpha_{21}+\beta_{21}) & p_{22}(\alpha_{21}+\beta_{21}) & \cdots & p_{S2}(\alpha_{21}+\beta_{21}) \\ \vdots & \vdots & \ddots & \vdots \\ p_{1S}(\alpha_{S1}+\beta_{S1}) & p_{2S}(\alpha_{S1}+\beta_{S1}) & \cdots & p_{SS}(\alpha_{S1}+\beta_{S1}) \end{pmatrix}.$$

A similar matrix has been introduced for the first time by Francq and Roussignol (1998) in the context of MS-GARCH models. See also Francq and Zakoian (2001) for Markov

Switching ARMA (MS-ARMA) models. Denote by $\rho(A)$ the spectral radius of the matrix A (i.e. the maximum of eigenvalues of A in modulus) and define the block-matrix Ω by

$$\Omega = \begin{pmatrix} M^{(1)} & M^{(2)} & \cdots & M^{(r-1)} & M^{(r)} \\ I_S & 0_{S\times S} & \cdots & 0_{S\times S} & 0_{S\times S} \\ 0_{S\times S} & I_S & \cdots & 0_{S\times S} & 0_{S\times S} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{S\times S} & 0_{S\times S} & \cdots & I_S & 0_{S\times S} \end{pmatrix}, \tag{3.1}$$

where $0_{S\times S}$ and I_S denote respectively the $S \times S$ null matrix and the $S \times S$ identity matrix. The following result gives a necessary and sufficient ergodicity condition for the MS-POLI model given by (2.6) and (2.7).

Theorem 3.1 *For $s = 1, ..., S$ let $F_\lambda = F_{s,\lambda}$ $(\lambda > 0)$ be a family of cdf's on $\mathcal{N}$ satisfying (2.3). There exists a stationary and ergodic sequence (Y_t) such that*

$$P\left(Y_t \leq y \mid \mathcal{F}_{t-1}^a\right) = F_{\Delta_t,\lambda_t}(y), \tag{3.2}$$

where λ_t satisfies (2.7) if

$$\rho(\Omega) < 1. \tag{3.3}$$

*Conversely, if there exists a solution of (3.2) such that $EY_t < \infty$ and the irreducible, stationary and ergodic Markov chain (Δ_t) satisfies **A1** then (3.3) holds.*

When $p = q = 1$ the condition (3.3) reduces to

$$\rho\left(M^{(1)}\right) < 1.$$

Under the latter condition, the unconditional mean of the process is given by

$$EY_t = \mathbf{1}^\top \left(I_S - M^{(1)}\right)^{-1} d \tag{3.4}$$

where $d = (\pi_1\omega_1, ..., \pi_S\omega_S)^\top$ and $\mathbf{1} = (1, ..., 1)^\top$ is a S-vector with unit components.

Remark 3.1 (POLI-X: POLI with exogenous variables). Consider the following MS-POLI-X extension of (2.7) with covariates

$$\lambda_t = \omega_{\Delta_t} + \sum_{i=1}^{q} \alpha_{\Delta_t,i} Y_{t-i} + \sum_{j=1}^{p} \beta_{\Delta_t,j} \lambda_{t-j} + \boldsymbol{\varphi}^\top \boldsymbol{X}_{t-1}, \tag{3.5}$$

where the components of the covariate $\boldsymbol{X}_t = (x_{1,t}, ..., x_{u,t})^\top$ are nonnegative numbers and $\boldsymbol{\varphi} = (\varphi_1, ..., \varphi_u)^\top \geq 0$ component-wise. If $\{(\boldsymbol{X}_t, \Delta_t)\}$ is stationary and ergodic then the conclusions of Theorem 3.1 hold true for model (3.5). This is not surprising since (3.5) replaces the constant intercept by a stationary intercept $\omega_t = \omega + \boldsymbol{\varphi}^\top \boldsymbol{X}_{t-1}$ and we have seen that the stationarity condition does not depend on the intercept. See also Francq and Thieu (2019) and Aknouche and Francq (2020).

Assume now the switching sequence (Δ_t) is iid with $\pi_s = P(\Delta_t = s)$. We call the model mixture POLI (M-POLI) in reference to mixture INGARCH models (Aknouche and Demmouche, 2019; Doukhan et *al*, 2018; Mao et *al*, 2019) in which (Δ_t) is assumed iid.

Remark 3.2 When (Δ_t) is iid and $p = q = 1$, the matrix Ω given by (3.1) reduces to

$$\Omega = M^{(1)} = \begin{pmatrix} \pi_1(\alpha_{11} + \beta_{11}) & \pi_1(\alpha_{11} + \beta_{11}) & \cdots & \pi_1(\alpha_{11} + \beta_{11}) \\ \pi_2(\alpha_{21} + \beta_{21}) & \pi_2(\alpha_{21} + \beta_{21}) & \cdots & \pi_2(\alpha_{21} + \beta_{21}) \\ \vdots & \vdots & \ddots & \vdots \\ \pi_S(\alpha_{S1} + \beta_{S1}) & \pi_S(\alpha_{S1} + \beta_{S1}) & \cdots & \pi_S(\alpha_{S1} + \beta_{S1}) \end{pmatrix}$$

with identical columns. Thus $\text{rank}(\Omega) = 1$ and by the rank-nullity theorem (e.g. Horn and Johnson, 2013, p. 6) the dimension of the kernel space corresponding to Ω is $S - 1$. Therefore, 0 is an eigenvalue with multiplicity $S - 1$, and the Sth eigenvalue is the trace $\sum_{s=1}^{S} \pi_s(\alpha_{s1} + \beta_{s1})$ of Ω which, by the non-negativity of the parameters, coincides with $\rho(\Omega)$. The condition (3.3) thus reduces to $\sum_{s=1}^{S} \pi_s(\alpha_{s1} + \beta_{s1}) < 1$. For general p and q the latter condition writes as follows.

Proposition 3.1 *For $s = 1, ..., S$ let $F_\lambda = F_{s,\lambda}$ $(\lambda > 0)$ be cdf's on $\mathcal{N}$ satisfying (2.3). There exists a stationary and ergodic sequence (Y_t) satisfying (3.2), where $\lambda_t := \lambda_{\Delta_t,t}$ satisfies (2.7) with (Δ_t) iid, if*

$$\sum_{s=1}^{S} \pi_s \left(\sum_{i=1}^{q} \alpha_{si} + \sum_{j=1}^{p} \beta_{sj} \right) < 1. \tag{3.6}$$

Conversely, if there exists a solution of (3.2) such that $EY_t = m < \infty$, where (Δ_t) is iid and is independent of $\mathcal{F}_{t-1}$, then (3.6) holds.

A proof of Proposition 3.1 is given in the supplement to this paper. Under (3.6) the mean of the process is given by

$$EY_t = \frac{\sum_{s=1}^{S} \pi_s \omega_s}{1 - \sum_{s=1}^{S} \pi_s \left(\sum_{i=1}^{q} \alpha_{si} + \sum_{j=1}^{p} \beta_{sj} \right)}. \tag{3.7}$$

Example 3.1 (Poisson MS-INGARCH(p,q)). Consider the following Poisson MS-INGARCH(p,q) model given by $Y_t \mid \mathcal{F}_{t-1}^a \sim \mathcal{P}(\lambda_{\Delta_t,t})$ where $\lambda_{\Delta_t,t}$ is given by (2.7) and $\mathcal{P}(\lambda)$ stands for the Poisson distribution with parameter λ. The Poisson distribution is a member of the class $\mathbb{F}$ given by (2.3). Thus a necessary and sufficient condition for the stationarity and ergodicity of (Y_t) is that (3.3) holds. If in particular (Δ_t) is iid the stationarity and ergodicity condition (3.3) reduces to (3.6). Such a condition is clearly sharper than the ones proposed by Aknouche and Demmouche (2019), Doukhan et *al* (2018) and Mao et *al* (2019) for the Poisson mixture INGARCH model. □

Example 3.2 (Negative Binomial MS-INGARCH(p,q)). A random variable Y follows a negative binomial, $Y \sim \mathcal{NB}(\upsilon, p)$, of parameters $\upsilon > 0$ and $p \in (0,1)$ if

$$P(Y = k) = \tfrac{\Gamma(k+\upsilon)}{k!\Gamma(\upsilon)} p^{\upsilon} (1-p)^k, \quad k \in \mathbb{N}.$$

(i) Assume that $Y_t \mid \mathcal{F}_{t-1}^a \sim \mathcal{NB}\left(\upsilon_{\Delta_t}, \frac{\upsilon_{\Delta_t}}{\upsilon_{\Delta_t}+\lambda_{\Delta_t,t}}\right)$ where $\upsilon_s > 0$ for $1 \le s \le S$ and $\lambda_{\Delta_t,t}$ is given by (2.7). We call this model NB2-MS-INGARCH. Since the latter distribution satisfies (2.3) (Aknouche and Francq, 2020), a necessary and sufficient condition for stationarity and ergodicity of (Y_t) is that (3.3) is satisfied. When (Δ_t) is iid the corresponding stationarity and ergodicity condition (3.6) is sharper than the condition provided by Aknouche and Demmouche (2019) for that model with $p = q = 1$.

(ii) Now let the NB1-MS-INGARCH model defined by $Y_t \mid \mathcal{F}_{t-1}^a \sim \mathcal{NB}\left(\upsilon_{\Delta_t}\lambda_{\Delta_t,t}, \frac{\upsilon_{\Delta_t}}{\upsilon_{\Delta_t}+1}\right)$. Aknouche and Francq (2020, Lemma 2.1) showed that such a distribution also satisfies (2.3). Therefore, a necessary and sufficient condition for stationarity and ergodicity of (Y_t) is given by (3.3) for the general Markov case and by (3.6) when (Δ_t) is iid □

Example 3.3 (Mixed MS-INGARCH). Let $S = 3$ and $Y_t \mid \mathcal{F}_{t-1} \sim F_{\Delta_t, \lambda_{\Delta_t,t}}$, where $\lambda_{\Delta_t,t}$ satisfies (2.7), $F_{1,\lambda_{1t}} \sim \mathcal{NB}\left(\upsilon_1\lambda_{1t}, \frac{\upsilon_1}{\upsilon_1+1}\right)$, $F_{2,\lambda_{2t}} \sim \mathcal{NB}\left(\upsilon_2, \frac{\upsilon_2}{\upsilon_2+\lambda_{2t}}\right)$ and $F_{3,\lambda_{3t}} \sim \mathcal{P}(\lambda_{3t})$ with $\upsilon_1, \upsilon_2 > 0$. Since the three mixing distributions belong to the class $\mathbb{F}$, a necessary and sufficient for the above model to be stationary and ergodic is that (3.3) holds true. □

Example 3.4 (Multiplicative MS-ACD). Aknouche and Francq (2020, Section 2.2) showed that the conditional distribution of any standard multiplicative ACD model of the form $Y_t = \lambda_t z_t$, where λ_t satisfies (2.3), belongs to the class of equal-stochastic mean order distributions given by (2.3). By the positive homogeneity of the quantile function $F^{-}_{\Delta_t,\lambda_{\Delta_t,t}}$ and in view of (2.3), the conditional distribution of any multiplicative MS-ACD of the form

$$Y_t = \lambda_{\Delta_t,t} z_t \tag{3.8}$$

with $\lambda_{\Delta_t,t}$ satisfying (2.7), also belongs to the class $\mathbb{F}$ in (2.3), where (z_t) is an iid sequence of positive random variables with $E(z_1) = 1$. Therefore, a necessary and sufficient condition for stationarity and ergodicity of (3.8) is given by (3.3) for a general irreducible stationary and ergodic Markov chain (Δ_t) satisfying **A1** and by (3.6) when (Δ_t) is iid. In particular, the latter result holds for the Gamma MS-ACD model given by $Y_t \mid \mathcal{F}^a_{t-1} \sim \Gamma(a_{\Delta_t}, a_{\Delta_t}\lambda_t)$, where λ_t satisfies (2.7), $a_s > 0$ for all $1 \leq s \leq S$, and $\Gamma(a, \lambda)$ stands for the Gamma distribution with shape parameter $a > 0$ and rate parameter $\lambda > 0$. $\square$

Remark 3.3 (Link with the ergodicity of multiplicative MS-ACD and MS-GARCH). It is easily seen that the square of the MS-GARCH process (X_t) of Francq and Zakoian (2005) is a multiplicative MS-ACD of the form $X_t^2 = \lambda_{\Delta_t,t} z_t$ where $\lambda_{\Delta_t,t}$ satisfies (2.7) and z_t is defined as in Example 3.4. Using the result of Example 3.4 the conditional distribution of Y_t^2 thus satisfies (2.3). Therefore, when Y_t in Theorem 3.1 is the square of an MS-GARCH whose squared volatility $\lambda_{\Delta_t,t}$ follows (2.7) we find the well known result that an MS-ACD is stationary with finite first-order moments (or a MS-GARCH is stationary with finite second-order moments) if and only if (3.3) holds true (e.g. Francq and Zakoian, 2008). $\square$

Turn now to the existence of moments for the MS-POLI(1,1) model. In Theorem 3.1 it has been shown that, provided the mixing distributions of the MS model satisfy the stochastic-equal-mean order property (2.3), the precise form of such distributions is not important for the strict stationarity and ergodicity. For the existence of moments, the next Theorem shows that the form of the conditional distribution influences indeed the ergodicity conditions. The same holds for the POLI model (2.1)-(2.3) (cf. Aknouche and Francq, 2020). For simplicity of notation, we write α_s and β_s instead of α_{s1} and β_{s1}.

Theorem 3.2 *Let $F_\lambda = F_{s,\lambda}$ $(1 \leq s \leq S,\ \lambda > 0)$ be a cdf on $\mathcal{N}$ satisfying (2.3) and $X_s \sim F_{s,\lambda}(x)$. Assume that, for some integer $\ell \geq 2$, there exist nonnegative coefficients $a_{sj}(0), a_{sj}(1), \ldots, a_{sj}(j)$ for all $j \leq \ell$ such that*

$$EX_s^j = \sum_{i=0}^{j} a_{sj}(i)\lambda^i, \quad j = 1, \ldots, \ell, \quad 1 \leq s \leq S. \tag{3.9}$$

*Under (3.3), let (Y_t) be a stationary sequence such that $P\left(Y_t \leq y \mid \mathcal{F}_{t-1}^a\right) = F_{\Delta_t,\lambda_t}(y)$, where $\lambda_t := \lambda_{\Delta_t,t}$ satisfies (2.7) with $p = q = 1$ and (Δ_t) fulfills **A1**. Then $EY_t^\ell < \infty$ if and only if*

$$\rho\left(M_\ell\right) < 1, \tag{3.10}$$

where M_ℓ is a $S \times S$-matrix defined by

$$M_\ell\left(s, \tau\right) = p_{\tau s} \sum_{j=0}^{\ell} a_s\left(j\right) \binom{\ell}{j} \alpha_s^j \beta_s^{\ell-j}, \quad s, \tau \in \{1, ...S\} \tag{3.11}$$

with $a_s(0) = a_s(1) = 1$ and $a_s(j) = a_{sj}(j)$ for $j \geq 2$ $(1 \leq s \leq S)$.

The matrix M_ℓ in (3.11) can be explicitly written as follows

$$M_\ell = \begin{pmatrix} p_{11} \sum\limits_{j=0}^{\ell} a_1\left(j\right) \binom{\ell}{j} \alpha_1^j \beta_1^{\ell-j} & p_{21} \sum\limits_{j=0}^{\ell} a_1\left(j\right) \binom{\ell}{j} \alpha_1^j \beta_1^{\ell-j} & \cdots & p_{S1} \sum\limits_{j=0}^{\ell} a_1\left(j\right) \binom{\ell}{j} \alpha_1^j \beta_1^{\ell-j} \\ p_{12} \sum\limits_{j=0}^{\ell} a_2\left(j\right) \binom{\ell}{j} \alpha_2^j \beta_2^{\ell-j} & p_{22} \sum\limits_{j=0}^{\ell} a_2\left(j\right) \binom{\ell}{j} \alpha_2^j \beta_2^{\ell-j} & \cdots & p_{S2} \sum\limits_{j=0}^{\ell} a_2\left(j\right) \binom{\ell}{j} \alpha_2^j \beta_2^{\ell-j} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1S} \sum\limits_{j=0}^{\ell} a_S\left(j\right) \binom{\ell}{j} \alpha_S^j \beta_S^{\ell-j} & p_{2S} \sum\limits_{j=0}^{\ell} a_S\left(j\right) \binom{\ell}{j} \alpha_S^j \beta_S^{\ell-j} & \cdots & p_{SS} \sum\limits_{j=0}^{\ell} a_S\left(j\right) \binom{\ell}{j} \alpha_S^j \beta_S^{\ell-j} \end{pmatrix}.$$

For $\ell = 1$, the latter matrix reduces to $M^{(1)}$ in (3.1). When (Δ_t) is iid, in view of Remark 3.2, the result of Theorem 3.2 reduces to the following.

Corollary 3.1 *For $s = 1, ..., S$ let $F_\lambda = F_{s,\lambda}$ $(\lambda > 0)$ be a cdf on $\mathcal{N}$ satisfying (2.3) and $X_s \sim F_{s,\lambda}(x)$. Assume that, for some integer $\ell \geq 2$, there exist nonnegative coefficients $a_{sj}(0), a_{sj}(1), \ldots, a_{sj}(j)$ for all $j \leq \ell$ such that EY_s^j has the form (3.10). Under (3.6), let (Y_t) be a stationary sequence such that $P\left(Y_t \leq y \mid \mathcal{F}_{t-1}^a\right) = F_{\Delta_t,\lambda_t}(y)$, where $\lambda_t := \lambda_{\Delta_t,t}$*

satisfies (2.7) *with* $p = q = 1$, *and* (Δ_t) *is iid with* $\pi_s = P(\Delta_t = s)$ *for all* $1 \leq s \leq S$. *We have* $EY_t^\ell < \infty$ *if and only if*

$$\sum_{s=1}^{S} \pi_s \sum_{j=0}^{\ell} a_s(j) \binom{\ell}{j} \alpha_s^j \beta_s^{\ell-j} < 1, \tag{3.12}$$

where $a_s(j)$ $(0 \leq j \leq \ell,\ 1 \leq s \leq S)$ *is given by* (3.12).

Example 3.1 (Continued) It is well known that if $Y_s \sim \mathcal{P}(\lambda_s)$ then $E\left(Y_s^\ell\right)$ satisfies (3.9) with $a_{sj}(i) = \left\{ {j \atop i} \right\}$ being the Stirling number of the second kind. Since $a_{sj}(j) = 1$ for all j, condition (3.10) becomes $\rho(M_\ell) < 1$ where

$$M_\ell = \begin{pmatrix} p_{11}(\alpha_1+\beta_1)^\ell & p_{21}(\alpha_1+\beta_1)^\ell & \cdots & p_{S1}(\alpha_1+\beta_1)^\ell \\ p_{12}(\alpha_2+\beta_2)^\ell & p_{22}(\alpha_2+\beta_2)^\ell & \cdots & p_{S2}(\alpha_2+\beta_2)^\ell \\ \vdots & \vdots & \ddots & \vdots \\ p_{1S}(\alpha_S+\beta_S)^\ell & p_{2S}(\alpha_S+\beta_S)^\ell & \cdots & p_{SS}(\alpha_S+\beta_S)^\ell \end{pmatrix}. \tag{3.13}$$

If in particular (Δ_t) is iid then the ℓ-th moment condition (3.12) reduces to

$$\sum_{s=1}^{S} \pi_s (\alpha_s+\beta_s)^\ell < 1. \ \square \tag{3.14}$$

Example 3.2 (Continued). (i) The first two moments of a random variable Y_s following the $\mathcal{NB}(\upsilon_s, \upsilon_s/(\lambda_s+\upsilon_s))$ distribution are given by $EY_s = \lambda_s$ and $EY_s^2 = \lambda_s + \frac{1+\upsilon_s}{\upsilon_s}\lambda_s^2$. Thus (3.9) holds with $a_s(2) = \frac{1+\upsilon_s}{\upsilon_s}$ $(1 \leq s \leq S)$ and the matrix M_ℓ for $\ell = 2$ writes

$$M_2 = \begin{pmatrix} p_{11}((\alpha_1+\beta_1)^2+\frac{\alpha_1^2}{\upsilon_1}) & p_{21}((\alpha_1+\beta_1)^2+\frac{\alpha_1^2}{\upsilon_1}) & \cdots & p_{S1}((\alpha_1+\beta_1)^2+\frac{\alpha_1^2}{\upsilon_1}) \\ p_{12}((\alpha_2+\beta_2)^2+\frac{\alpha_2^2}{\upsilon_2}) & p_{22}((\alpha_2+\beta_2)^2+\frac{\alpha_2^2}{\upsilon_2}) & \cdots & p_{S2}((\alpha_2+\beta_2)^2+\frac{\alpha_2^2}{\upsilon_2}) \\ \vdots & \vdots & \ddots & \vdots \\ p_{1S}((\alpha_S+\beta_S)^2+\frac{\alpha_S^2}{\upsilon_S}) & p_{2S}((\alpha_S+\beta_S)^2+\frac{\alpha_S^2}{\upsilon_S}) & \cdots & p_{SS}((\alpha_S+\beta_S)^2+\frac{\alpha_S^2}{\upsilon_S}) \end{pmatrix}.$$

In particular, when (Δ_t) is iid the moment condition (3.12) reduces to

$$\sum_{s=1}^{S} \pi_s((\alpha_s+\beta_s)^2 + \frac{\alpha_s^2}{\upsilon_s}) < 1.$$

Conditions for the existence of the third and fourth moments are obtained accordingly; see Aknouche and Francq (2020) for the POLI model corresponding to $S = 1$.

(ii) It is easily seen that the moments EY_s^ℓ of the $\mathcal{NB}(\upsilon_s\lambda_s, \upsilon_s/(1+\upsilon_s))$ distribution satisfy (3.10) with $a_s(j) = a_{sj}(j) = 1$ for all j and s (Aknouche and Francq, 2020). Therefore, the necessary and sufficient condition for the NB1-MS-INGARCH model is given by (3.13) for the general Markov case and by (3.14) when (Δ_t) is iid. $\square$

Example 3.3 (Continued). For Example 3.3, the matrix M_ℓ for $\ell = 2$ is given by

$$M_2 = \begin{pmatrix} p_{11}\left(\alpha_1+\beta_1\right)^2 & p_{21}\left(\alpha_1+\beta_1\right)^2 & p_{31}\left(\alpha_1+\beta_1\right)^2 \\ p_{12}((\alpha_2+\beta_2)^2+\frac{\alpha_2^2}{\upsilon_2}) & p_{22}((\alpha_2+\beta_2)^2+\frac{\alpha_2^2}{\upsilon_2}) & p_{32}((\alpha_2+\beta_2)^2+\frac{\alpha_2^2}{\upsilon_2}) \\ p_{13}\left(\alpha_3+\beta_3\right)^2 & p_{23}\left(\alpha_3+\beta_3\right)^2 & p_{33}\left(\alpha_3+\beta_3\right)^2 \end{pmatrix}.$$

$\square$

3.2 Extension to nonlinear conditional mean forms

We extend Theorem 3.1 to the case where the conditional mean λ_t satisfies the more general specification (2.8). Assume that the functions $g_s(y_1,\ldots,y_q,\lambda_1,\ldots,\lambda_p)$ $(1\leq s\leq S)$ in (2.8) are Lipschitz in the sense that, for all (y_i, y_i'), $i=1,\ldots,q$ and for all $\left(\lambda_j,\lambda_j'\right)$, $j=1,\ldots,p$, there are positive constants α_{si} and β_{sj} such that

$$\left|g_s(y_1,\ldots,y_q,\lambda_1,\ldots,\lambda_p)-g_s(y_1',\ldots,y_q',\lambda_1',\ldots,\lambda_p')\right| \leq \sum_{i=1}^{q}\alpha_{si}\left|y_i-y_i'\right|+\sum_{j=1}^{p}\beta_{sj}|\lambda_j-\lambda_j'|. \tag{3.15}$$

Theorem 3.3 *For $s=1,...,S$ let $F_\lambda = F_{s,\lambda}$ $(\lambda>0)$ be a family of cdf's on $\mathcal{N}$ satisfying (2.3) and $g_s(y_1,\ldots,y_q,\lambda_1,\ldots,\lambda_p)$ $(1\leq s\leq S)$ be S Lipschitz functions satisfying (3.15). Let the matrix Ω be defined as in (3.1) where the coefficients (α_{si},β_{sj}) are those in the right-hand-side of the Lipschitz inequality (3.15). There exists a stationary and ergodic sequence (Y_t) such that the distribution of Y_t conditional on $\mathcal{F}_{t-1}^a$ is F_{Δ_t,λ_t}, where λ_t satisfies (2.8) with (Δ_t) a stationary and ergodic Markov chain, if $\rho(\Omega)<1$.*

If the mixture sequence (Δ_t) is iid with $\pi_s = P(\Delta_t = s)$ then Theorem 3.1 becomes.

Proposition 3.2 *For $s=1,...,S$ let $F_\lambda = F_{s,\lambda}$ $(\lambda>0)$ be a family of cdf's on $\mathcal{N}$ satisfying (2.3). Assume that the functions $g_s(y_1,\ldots,y_q,\lambda_1,\ldots,\lambda_p)$ $(1\leq s\leq S)$ satisfy (3.15)* *and that* (Δ_t) *is iid. If*

$$\sum_{s=1}^{S}\pi_s\left(\sum_{i=1}^{q}\alpha_{si}+\sum_{j=1}^{p}\beta_{sj}\right)<1 \tag{3.16}$$

then there exists a stationary and ergodic sequence (Y_t) *such that the distribution of* Y_t *conditional on* $\mathcal{F}_{t-1}^{a}$ *is* F_{Δ_t,λ_t}*, where* λ_t *satisfies* (2.8)*.*

The proof of Proposition 3.2 is given in the supplement to this paper.

Remark 3.4 (Exogenous variables) If we consider the following specification

$$\lambda_t = g_{\Delta_t}(Y_{t-1},\ldots,Y_{t-q},\lambda_{t-1},\ldots,\lambda_{t-p})+h(\boldsymbol{X}_{t-1})$$

instead of (2.8), where h is a positive function and $\boldsymbol{X}_{t-1}$ is defined as in Remark 3.1, then the conclusions of Theorem 3.3 remains unchanged. Here again the integration of covariates in the general specification of the conditional mean, as an additive function, does not influence the stability conditions.

4 The present-regime dependent switching

In this Section we give stationarity and ergodicity conditions for model (2.6) when the specification of $\lambda_{\Delta_t,t}$ is governed by (2.9) or (2.10). We first consider the linear conditional mean case (2.9). Particular equivalent conditions when (Δ_t) is iid are obtained. We then extend the result to the more general specification (2.10).

4.1 The linear conditional mean case

Denote by $diag\,(a_1,...,a_S)$ the diagonal matrix whose diagonal elements are $a_1,...,a_S$ in this order. Set $\alpha_i=(\alpha_{1i},...,\alpha_{Si})^{\top}$, $(1\le i\le q)$, $\boldsymbol{\beta}_j = diag\left(\beta_{1j},...,\beta_{Sj}\right)$, $(1\le j\le p)$ and let the S-vector $\mathbf{1}_s$ defined by $\mathbf{1}_s(\tau)=\begin{cases}1 & \text{if } \tau=s\\ 0 & \text{otherwise}\end{cases}$ $(1\le\tau\le S)$. Consider the $S^2\times S^2$-matrix

$D^{(l)}$ $(l = 1, ..., r = \max(p, q))$ given by

$$D^{(l)} = \begin{pmatrix} D_{11}^{(l)} & \cdots & D_{1S}^{(l)} \\ \vdots & \ddots & \vdots \\ D_{S1}^{(l)} & \cdots & D_{SS}^{(l)} \end{pmatrix} \text{ and } D_{\tau s}^{(l)} = p_{s\tau}^{(l)} \left(\alpha_l \mathbf{1}_s^\top + \boldsymbol{\beta}_l \right), \quad \tau, s = 1, ..., S.$$

Let also

$$D = \begin{pmatrix} D^{(1)} & D^{(2)} & \cdots & D^{(r-1)} & D^{(r)} \\ I_{S^2} & 0_{S^2 \times S^2} & \cdots & 0_{S^2 \times S^2} & 0_{S^2 \times S^2} \\ 0_{S^2 \times S^2} & I_{S^2} & \cdots & 0_{S^2 \times S^2} & 0_{S^2 \times S^2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{S^2 \times S^2} & 0_{S^2 \times S^2} & \cdots & I_{S^2} & 0_{S^2 \times S^2} \end{pmatrix}. \tag{4.1}$$

A similar matrix has been earlier proposed by Abramson and Cohen (2007) in the real-valued MS-GARCH case.

Theorem 4.1 *For $s = 1, ..., S$ let $F_\lambda = F_{s,\lambda}$ $(\lambda > 0)$ be a family of cdf's on $\mathcal{N}$ satisfying (2.3) and assume that (Δ_t) in (2.6) is a stationary and ergodic Markov chain. There exists a stationary and ergodic sequence (Y_t) satisfying (2.6), where the λ_{st}'s satisfy (2.9), if*

$$\rho(D) < 1. \tag{4.2}$$

*Conversely, if there exists a solution of (2.6) and (2.9) such that $EY_t < \infty$ and (Δ_t) is irreducible and satisfies **A1** then (4.2) holds.*

For example, when $p = q = 1$ and $S = 2$, condition (4.2) reduces to $\rho\left(D^{(1)} < 1\right)$ where

$$D^{(1)} = \begin{pmatrix} p_{11}(\alpha_{11} + \beta_{11}) & 0 & p_{21}\beta_{11} & p_{21}\alpha_{11} \\ p_{11}\alpha_{21} & p_{11}\beta_{21} & 0 & p_{21}(\alpha_{21} + \beta_{21}) \\ p_{12}(\alpha_{11} + \beta_{11}) & 0 & p_{22}\beta_{11} & p_{22}\alpha_{11} \\ p_{12}\alpha_{21} & p_{12}\beta_{21} & 0 & p_{22}(\alpha_{21} + \beta_{21}) \end{pmatrix}.$$

Let $\otimes$ be the Kronecker product and denote by $Vec(A)$ the stacking vector operator for a matrix A. Under $\rho\left(D^{(1)}\right) < 1$ the mean of the stationary solution is given by

$$EY_t = Vec(I_S)^\top \left(I_{S^2} - D^{(1)}\right)^{-1} \pi \otimes \omega \tag{4.3}$$

where $\omega = (\omega_1, ..., \omega_S)^\top$ and $\pi = (\pi_1, ..., \pi_S)^\top$.

Remark 4.1 When $p = q = 1$, it may be possible to use a representation similar to that of Liu (2006) for the MS-GARCH$(1,1)$ model. Let (U_t) be an iid sequence of random variables uniformly distributed in $[0,1]$, independent of the sequence (Δ_t). Define the S-vectors $\boldsymbol{\lambda}_t = (\lambda_{1t}, ..., \lambda_{St})^\top$, $\mathbf{1}_{[\Delta_t]} = \left(1_{[\Delta_t=1]}, ..., 1_{[\Delta_t=S]}\right)^\top$ and $\mathbf{F}^-_{\boldsymbol{\lambda}_t} = \left(F^-_{1,\lambda_{1t}}(U_t), ..., F^-_{S,\lambda_{St}}(U_t)\right)^\top$. Assume that

$$Y_t = \sum_{s=1}^{S} 1_{[\Delta_t=s]} F^-_{s,\lambda_{st}}(U_t) \tag{4.4}$$

which is a particular form of (2.6), where λ_{st} satisfies (2.9) with $p = q = 1$. Then model (4.4) can be written in the following vector form

$$\boldsymbol{\lambda}_t = \omega + \alpha_1 \mathbf{1}^\top_{[\Delta_{t-1}]} \mathbf{F}^-_{\boldsymbol{\lambda}_{t-1}} + \boldsymbol{\beta}_1 \boldsymbol{\lambda}_{t-1},$$

where α_1 and $\boldsymbol{\beta}_1$ are defined as in (4.1). Hence, by **A1** and Lemma 3 in Francq and Zakoian (2005), we get

$$\pi_s E(\boldsymbol{\lambda}_t \mid \Delta_{t-1} = s) = \pi_s \omega + \left(\alpha_1 \mathbf{1}^\top_s + \boldsymbol{\beta}_1\right) \sum_{\tau=1}^{S} p_{\tau s} \pi_\tau E(\boldsymbol{\lambda}_{t-1} \mid \Delta_{t-2} = \tau).$$

In view of the latter equality and using similar techniques as above, a sufficient condition for the model given by (4.4) and (2.9) with $p = q = 1$ to have a stationary and ergodic solution is that

$$\rho(C) < 1,$$

where

$$C = \begin{pmatrix} C_{11} & \cdots & C_{S1} \\ \vdots & \ddots & \vdots \\ C_{1S} & \cdots & C_{SS} \end{pmatrix} \text{ and } C_{\tau s} = p_{\tau s}\left(\alpha_1 \mathbf{1}^\top_s + \boldsymbol{\beta}_1\right), \quad \tau, s = 1, ..., S.$$

For $S = 2$ the above matrix is explicitly given by

$$C = \begin{pmatrix} p_{11}(\alpha_{11} + \beta_{11}) & 0 & p_{21}(\alpha_{11} + \beta_{11}) & 0 \\ p_{11}\alpha_{21} & p_{11}\beta_{21} & p_{21}\alpha_{21} & p_{21}\beta_{21} \\ p_{12}\beta_{11} & p_{12}\alpha_{11} & p_{22}\beta_{11} & p_{22}\alpha_{11} \\ 0 & p_{12}(\alpha_{21} + \beta_{21}) & 0 & p_{22}(\alpha_{21} + \beta_{21}) \end{pmatrix}.$$

Under the condition $\rho(C) < 1$, the unconditional mean of the stationary solution is given by

$$EY_t = Vec(I_S)^\top (I_{S^2} - C)^{-1} \pi \otimes \omega. \tag{4.5}$$

For the iid switching case we use an equivalent but simpler representation than (4.1). Assume now that (Δ_t) is iid. In the INGARCH case, a similar model was considered by Diop et *al* (2016) when the conditional distribution is a Poisson mixture and by Diop et *al* (2018) for a negative binomial mixture. These authors studied stationarity in mean of the models but without considering strict stationarity and ergodicity. Let

$$\begin{aligned} A^{(l)} &= \alpha_l \pi^\top + \boldsymbol{\beta}_l \\ &= \begin{pmatrix} \alpha_{1l}\pi_1 + \beta_{1l} & \alpha_{1l}\pi_2 & \cdots & \alpha_{1l}\pi_S \\ \alpha_{2l}\pi_1 & \alpha_{2l}\pi_2 + \beta_{2l} & \cdots & \alpha_{2l}\pi_S \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{Sl}\pi_1 & \alpha_{Sl}\pi_2 & \cdots & \alpha_{Sl}\pi_S + \beta_{Sl} \end{pmatrix}, \quad 1 \le l \le r, \end{aligned}$$

and

$$\Sigma = \begin{pmatrix} A^{(1)} & A^{(2)} & \cdots & A^{(r-1)} & A^{(r)} \\ I_S & 0_{S\times S} & \cdots & 0_{S\times S} & 0_{S\times S} \\ 0_{S\times S} & I_S & \cdots & 0_{S\times S} & 0_{S\times S} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{S\times S} & 0_{S\times S} & \cdots & I_S & 0_{S\times S} \end{pmatrix}. \tag{4.6}$$

Theorem 4.2 *For $s = 1, ..., S$ let $F_\lambda = F_{s,\lambda}$ $(\lambda > 0)$ be a family of cdf's on $\mathcal{N}$ satisfying (2.3). A sufficient condition for the existence of a stationary and ergodic sequence (Y_t) with conditional distribution of the form (2.6), where λ_{st} satisfies (2.9) and (Δ_t)* is *iid, is that*

$$\rho(\Sigma) < 1. \tag{4.7}$$

Conversely, if there exists a solution to (2.6) and (2.9) *such that $EY_t < \infty$ and (Δ_t) is iid with Δ_t is independent of $(\lambda_{st})_s$ then* (4.7) *holds.*

For the particular case $p = q = 1$ the matrix Σ given by (4.6) reduces to

$$A^{(1)} = \begin{pmatrix} \alpha_{11}\pi_1 + \beta_{11} & \alpha_{11}\pi_2 & \cdots & \alpha_{11}\pi_S \\ \alpha_{21}\pi_1 & \alpha_{21}\pi_2 + \beta_{21} & \cdots & \alpha_{21}\pi_S \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{S1}\pi_1 & \alpha_{S1}\pi_2 & \cdots & \alpha_{S1}\pi_S + \beta_{S1} \end{pmatrix}.$$

Under (4.7) with $p = q = 1$ the mean of the process has the following expression

$$EY_t = \pi^\top \left(I_S - A^{(1)}\right)^{-1} \omega. \tag{4.8}$$

4.2 Nonlinear conditional mean

Theorem 4.1 can be extended to the case where the mixing conditional means $(\lambda_{st})_s$ have nonlinear specifications of the form (2.10). Berentsen et *al* (2018) proposed a Markov-switching INGARCH model, a particular case of (2.10), but they do not study its ergodic properties. Theorem 4.3 below give an answer to that question. Let the matrix D be defined as in (4.1) while replacing the coefficients $(\alpha_{si}, \beta_{sj})$ by those of the Lipschitz inequality (3.15).

Theorem 4.3 *For $s = 1, ..., S$ let $F_\lambda = F_{s,\lambda}$ $(\lambda > 0)$ be a family of cdf's on $\mathcal{N}$ satisfying (2.3). Assume that the functions $g_s(y_1, \ldots, y_q, \lambda_1, \ldots, \lambda_p)$ $(1 \leq s \leq S)$ satisfy the Lipschitz condition (3.15). If $\rho(D) < 1$ then there exists a stationary and ergodic sequence (Y_t) such that the distribution of Y_t conditional on $\mathcal{F}^a_{t-1}$ is $F_{\Delta_t, \lambda_{\Delta_t, t}}(y)$, where λ_{st} satisfies (2.10) and (Δ_t) in (2.6) is a stationary and ergodic Markov chain.*

When the regime sequence (Δ_t) is iid, the following result simplify the condition of the latter theorem and at the same time generalizes Theorem 4.2 to the case where the specification of the conditional mean is given by the general form (2.10).

Theorem 4.4 *For $s = 1, ..., S$ let $F_\lambda = F_{s,\lambda}$ $(\lambda > 0)$ be a family of cdf's on $\mathcal{N}$ satisfying (2.3). Assume that the functions $g_s(y_1, \ldots, y_q, \lambda_1, \ldots, \lambda_p)$ $(1 \leq s \leq S)$ satisfy the Lipschitz condition (3.15). Let Σ be defined as in (4.6) but whose coefficients are those of the Lipschitz representation (3.15). If $\rho(\Sigma) < 1$ then there exists a stationary and ergodic sequence (Y_t) such that the distribution of Y_t conditional on $\mathcal{F}^a_{t-1}$ is $F_{\Delta_t, \lambda_{\Delta_t, t}}(y)$, where λ_{st} satisfies (2.10) with (Δ_t) iid.*

The proofs of Theorem 4.2 and Theorem 4.4 are available in the supplement to this paper.

5 The present-regime mean-dependent switching

This Section deals with the MS-POLI model (2.6) with mixing conditional means satisfying (2.12). The mean stationarity of an MS-GARCH analog to (2.6) and (2.12) was studied by Klaassen (2002) when $p = q = 1$ while the strict stationarity and ergodicity were investigated by Abramson and Cohen (2007). We thus give a necessary and sufficient ergodicity condition for (2.6) and (2.12) provided the corresponding mixing distributions satisfy the stochastic-equal-mean property (2.3). Ergodicity conditions for the extension (2.13) are also provided. Let Ω be the matrix defined by (3.1).

Theorem 5.1 *For $s = 1, ..., S$ let $F_\lambda = F_{s,\lambda}$ $(\lambda > 0)$ be a family of cdf's on $\mathcal{N}$ satisfying (2.3) and assume that (Δ_t) is a stationary and ergodic Markov chain. There exists a stationary and ergodic sequence (Y_t) satisfying (2.6), where λ_{st} is given by (2.12), if (3.3) is satisfied.*

*Conversely, if there exists a solution of (2.6) and (2.12) such that $EY_t < \infty$ with (Δ_t) an irreducible stationary Markov chain satisfying **A1**, then (3.3) holds.*

When $p = q = 1$, the ergodicity condition of model (2.6) and (2.12) is $\rho\left(M^{(1)}\right) < 1$ where $M^{(1)}$ is given by (3.1). Under the latter condition, the mean of the stationary solution is given by

$$EY_t = \mathbf{1}^\top \left(I_S - M^{(1)}\right)^{-1} d.$$

The latter result is the same unconditional mean formula given for the past-dependent regime case. It is also similar to the unconditional variance expression for the MS-GARCH(1,1) model of Klaassen (2002, equalities (25)-(27)).

Let us finally generalize Theorem 5.1 to the case where the conditional mean regimes are given by the general nonlinear form (2.13).

Theorem 5.2 *For $s = 1, ..., S$ let $F_\lambda = F_{s,\lambda}$ $(\lambda > 0)$ be a family of cdf's on $\mathcal{N}$ satisfying (2.6). Assume that the functions $g_s(y_1, \ldots, y_q, \mu_1, \ldots, \mu_p)$ $(1 \le s \le S)$ in (2.13) satisfy the*

Lipschitz condition (3.15)*. Let* Ω *be defined as in* (3.1) *but whose coefficients are those of the Lipschitz representation* (3.15)*. If* (3.3) *is satisfied then there exists a stationary and ergodic sequence* (Y_t) *such that the distribution of* Y_t *conditional on* $\mathcal{F}^a_{t-1}$ *is* $F_{\Delta_t,\lambda_{\Delta_t,t}}$*, where* λ_{st} *satisfies* (2.13)*.*

6 Monte Carlo approximations of expectations

The aims of this short section are to: 1) compare the theoretical expectations of the different MS-POLI models with their empirical counterparts; 2) check numerically that the different expressions of a same expectation (such as (4.3) and (4.5)) coincide; 3) see the effect of the regime switching mechanism on the marginal expectation. Note that 1) will serve as an indirect validation of our theoretical calculations. Note also that 2) is useful because, in particular, we have not been able to directly prove that (4.3) and (4.5) provide the same value although the matrices $D^{(1)}$ and C are not the same.

For the conditional distribution (2.6) we first consider a continuous ACD type model $Y_t = \lambda_{\Delta_t,t} z_t$ where (z_t) is iid with exponential distribution of mean 1, and is independent of (Δ_t). We also consider an integer-valued conditional distribution, more precisely a negative-Binomial $Y_t \mid \mathcal{F}^a_{t-1} \sim \mathcal{NB}\left(1, \frac{1}{1+\lambda_{\Delta_t,t}}\right)$. For the distribution of (Δ_t) we also consider two cases. First a Markov switching with 3 regimes and transition probability matrix

$$P = \begin{pmatrix} 0.1 & 0.1 & 0.8 \\ 0.1 & 0.1 & 0.8 \\ 0.8 & 0.1 & 0.1 \end{pmatrix} \text{ (so that } \pi_1 = 0.429..., \ \pi_2 = 0.1, \ \pi_3 = 0.470...),$$

and secondly an iid sequence with $\pi_1 = \pi_2 = 0.1$ and $\pi_3 = 0.8$. For convenience we refer to (2.7), (2.9) and (2.12) as Type I, II, and III regime switching, respectively. The other parameters are fixed to $\omega = (1, 2, 3)$, $p = q = 1$, $\alpha_1 = (0.05, 0.1, 0.1)$ and $\beta_1 = (0.5, 0.6, 0.7)$. We end up with a set of 12 models. For example, NB-MS-II denotes an INGARCH model, with conditional negative-Binomial distribution of mean satisfying the present-regime mean-dependent switching (2.9), in which the Markov chain Δ_t has transition probability matrix

Table 1: Marginal mean of the model when the Markov chain Δ_t is independent (IID) or not (MS), and as function of the regime switching type (ACD and NB, as well as Type I and III, have same mean)

MS Type I	MS Type II	IID Type I	IID Type II
6.137017	7.764563	11.48936	12.45823

P. Table 1 provides the theoretical means of these models, obtained from (3.4) for Type I or Type III and from (4.3) for the regime switching of Type II. We verified that (4.3) and (4.5) always give the same result, and that (3.4) and (3.7), and (4.3) and (4.8), coincide when (Δ_t) is iid. Note also that, as expected, the mean of the process depends of its regime switching mechanism. We then simulated $N = 100$ independent replications of simulations of length $n = 1000$ of each of the 12 models. Figure 1 shows that the empirical means are in accordance with the theoretical ones. Similar results were obtained for other sets of parameters, which leads us to consider that our theoretical calculations are plausible.

7 Conclusion

We proposed stationarity and ergodicity conditions for a vast and flexible class of Markov switching autoregressive conditional mean models. Such a class includes in particular count, duration and proportion models without imposing the MEM constraint while having a similar dynamics for the conditional mean as the MS-GARCH models. The switching sequence can be iid or a stationary and ergodic Markov chain. In addition, the mixing distributions are not constrained to be equal across regimes but should belong to the class of equal stochastic and mean orders, which includes the one-parameter exponential family and other interesting distributions. Three variants of the model differing in the way the regime sequence is integrated in the conditional mean dynamics have been considered. We also proposed for the case $p = q = 1$ conditions for finite moments when the conditional mean specification is governed by the past of the regime sequence. The stationarity conditions match those

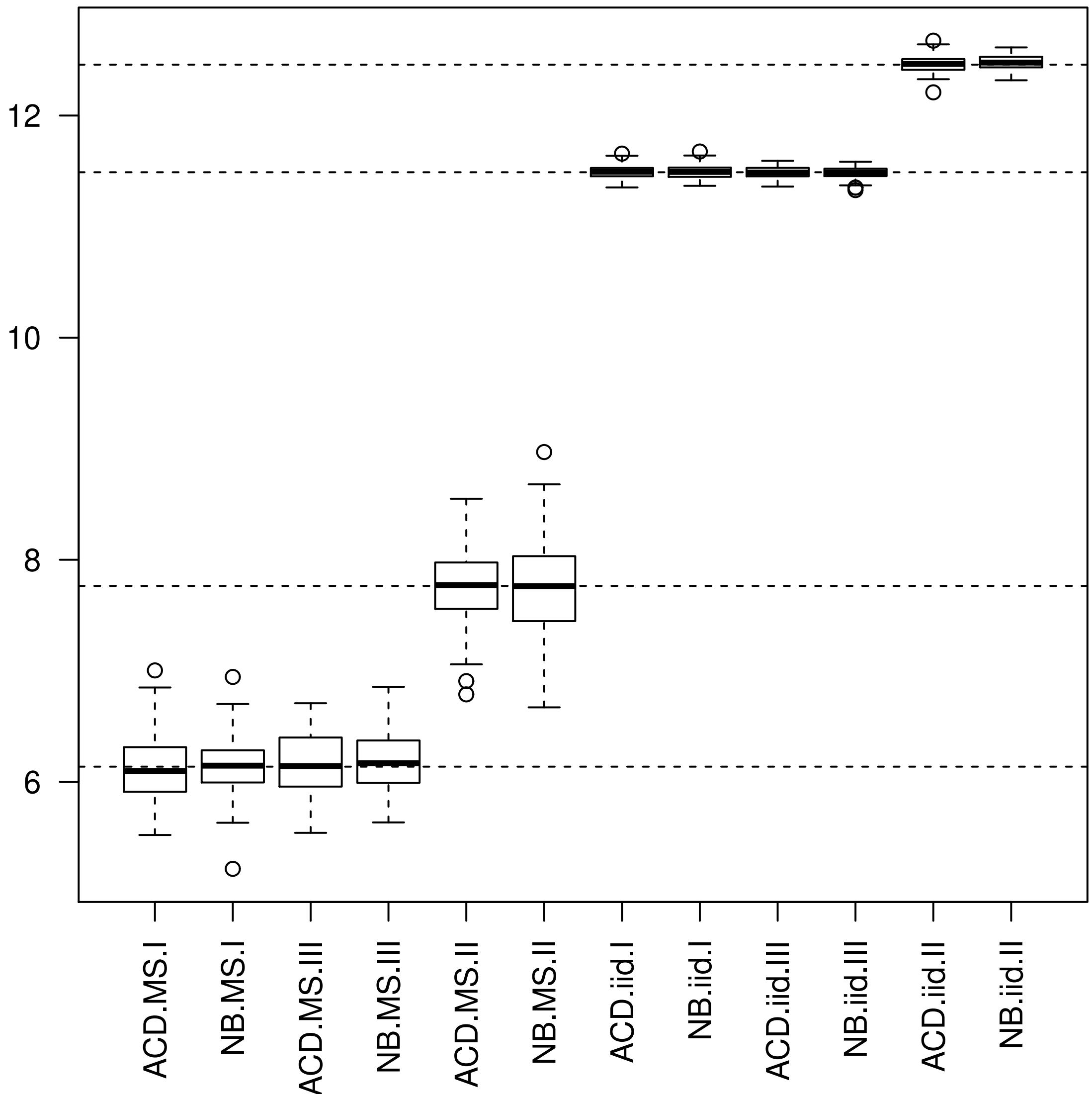

Figure 1: Boxplots of $N = 100$ empirical means of simulation of length $n = 1000$ of 12 MS-POLI models (the theoretical means of Table 1 are indicated by dotted lines).

known for similar MS-GARCH models irrespective of the choice of mixing distributions. However, finite moment conditions heavily depend on these mixing distributions. Numerous extensions can be considered. In particular, it would be interesting to extend our results to multivariate versions of the MS-POLI model.

A Proofs

Lemma A.1 *Let A be a (component-wise) nonnegative $r \times r$-matrix ($r \in \{1, 2, ...\}$), and b a positive vector in $\mathbb{R}^r$. If there is a positive vector $c \in \mathbb{R}^r$ such that*

$$b = Ab + c$$

then $\rho(A) < 1$.

Proof Under the assumptions of the lemma, it follows that

$$Ab < b$$

(component-wise) so the result follows from Corollary 8.1.29 of Horn and Johnson (2013). □

Proof of Theorem 3.1 If there exists a solution of (3.2) such that $EY_t = E\lambda_t < \infty$ for all t, then for all $1 \leq s \leq S$

$$\begin{aligned} E\left(Y_t \mid \Delta_t = s\right) &= E\left(E\left(Y_t \mid \mathcal{F}_{t-1}^a\right) \mid \Delta_t = s\right) \\ &= E\left(\lambda_t \mid \Delta_t = s\right). \end{aligned}$$

Under **A1**, using Lemma 3 in Francq and Zakoian (2005) we thus get

$$\begin{aligned} \pi_s E\left(\lambda_t \mid \Delta_t = s\right) &= \pi_s \omega_s + \pi_s \sum_{l=1}^{r} \left(\alpha_{sl} + \beta_{sl}\right) E\left(\lambda_{t-l} \mid \Delta_t = s\right), \; 1 \leq s \leq S \\ &= \pi_s \omega_s + \sum_{l=1}^{r} \sum_{j=1}^{S} \left(\alpha_{sl} + \beta_{sl}\right) p_{js}^{(l)} \pi_j E\left(\lambda_{t-l} \mid \Delta_{t-l} = j\right). \end{aligned} \tag{A.1}$$

Taking $h_s = \pi_s E\left(\lambda_t \mid \Delta_t = s\right)$, $h = (h_1, ..., h_S)^\top$, $\mathbf{h} = \left(h^\top, ..., h^\top\right)_{rS\times 1}^\top$, $d = (\pi_1\omega_1, ..., \pi_S\omega_S)^\top$, and $\mathbf{d} = \left(d^\top, 0_{S\times 1}^\top, ..., 0_{S\times 1}^\top\right)_{rS\times 1}^\top$, equality (A.1) can be rewritten in a block-matrix form $\mathbf{h} = \mathbf{d} + \Omega\mathbf{h}$, which entails

$$\mathbf{h}_S = \mathbf{d}_S + \Omega\mathbf{h}_S \quad \text{with} \quad \mathbf{h}_S = \Omega^{S-1}\mathbf{h}, \quad \mathbf{d}_S = \Omega^{S-1}\mathbf{d}.$$

Therefore, by the positivity of the coefficients of $\mathbf{h}_S$ and $\mathbf{d}_S$ and Lemma A.1, condition (3.3) should be satisfied.

Conversely, assume (3.2) holds. Let (U_t) be an iid sequence of random variables uniformly distributed in $[0, 1]$, independent of the sequence (Δ_t). Let also $Y_t^{(k)} = \lambda_t^{(k)} = 0$ for $k < 0$, and when $k \geq 0$,

$$Y_t^{(k)} = F^-_{\Delta_t, \lambda_t^{(k)}}(U_t), \qquad \lambda_t^{(k)} = \omega_{\Delta_t} + \sum_{i=1}^{q} \alpha_{\Delta_t, i} Y_{t-i}^{(k-i)} + \sum_{j=1}^{p} \beta_{\Delta_t, j} \lambda_{t-j}^{(k-j)}. \tag{A.2}$$

For $k \geq 2$, we have

$$\lambda_t^{(k)} = \psi_k(U_{t-1}, \ldots, U_{t-k+1}; \Delta_t, ..., \Delta_{t-k+1}),$$

where $\psi_k : [0, 1]^{k-1} \times \{1, ..., S\}^k \to [0, \infty)$ is a measurable function. Therefore, for any k, the sequences $\left(\lambda_t^{(k)}\right)_t$ and $\left(Y_t^{(k)}\right)_t$ are stationary and ergodic. Let also $\mathcal{F}_{t-1}^{a(k)}$ and $\mathcal{F}_{t-1}^*$ be the sigma-fields generated by $\left\{Y_{t-i}^{(k-i)}, \Delta_{t-i+1}, i > 0\right\}$ and $\{U_s, \Delta_{s+1}, s < t\}$, respectively. We have

$$E\left(Y_t^{(k)} \mid \mathcal{F}_{t-1}^{a(k)}\right) = E\left(Y_t^{(k)} \mid \mathcal{F}_{t-1}^*\right) = \lambda_t^{(k)},$$
$$P\left(Y_t^{(k)} \leq y \mid \mathcal{F}_{t-1}^{a(k)}\right) = P\left(F^-_{\Delta_t, \lambda_t^{(k)}}(U_t) \leq y \mid \mathcal{F}_{t-1}^*\right) = F_{\Delta_t, \lambda_t^{(k)}}(y).$$

To show the existence of a solution to (3.2), with $\mathcal{F}_{t-1}^a$ replaced by $\mathcal{F}_{t-1}^*$, it is sufficient to show that

$$\lambda_t = \lim_{k \to \infty} \lambda_t^{(k)} \text{ exists almost surely (a.s.) in } [0, +\infty). \tag{A.3}$$

Taking the limit as $k \to \infty$ in both sides of the equalities in (A.2), the solution will be then given by $Y_t = \lim_{k\to\infty} Y_t^{(k)} = F^-_{\Delta_t, \lambda_t}(U_t)$ a.s. Note that the distribution of Y_t given $\mathcal{F}_{t-1}^*$ is the same as that of Y_t given $\mathcal{F}_{t-1}^a$ since λ_t is $\mathcal{F}_{t-1}^a$-measurable.

We now show (A.3) under (A.2). We first prove that, for all positive integer k,

$$0 \leq \lambda_t^{(k-1)} \leq \lambda_t^{(k)} \quad \text{a.s.} \tag{A.4}$$

and

$$E\left(Y_t^{(k)} - Y_t^{(k-1)}\right) = E\left(\lambda_t^{(k)} - \lambda_t^{(k-1)}\right) \in [0, \infty). \tag{A.5}$$

When $k \leq 0$, it is clear that (A.4) and (A.5) hold true. Assume (A.4) is satisfied up to k. In view of (2.3)

$$\begin{aligned}
\lambda_t^{(k)} &= \omega_{\Delta_t} + \sum_{i=1}^{q} \alpha_{\Delta_t,i} F^{-}_{\Delta_{t-i},\lambda_{t-i}^{(k-i)}}(U_{t-i}) + \sum_{j=1}^{p} \beta_{\Delta_t,j} \lambda_{t-j}^{(k-j)} \\
&\leq \omega_{\Delta_t} + \sum_{i=1}^{q} \alpha_{\Delta_t,i} F^{-}_{\Delta_{t-i},\lambda_{t-i}^{(k+1-i)}}(U_{t-i}) + \sum_{j=1}^{p} \beta_{\Delta_t,j} \lambda_{t-j}^{(k+1-j)} = \lambda_t^{(k+1)}.
\end{aligned}$$

Therefore (A.4) and (A.5) follow by induction. Note that $EY_t^{(k)} = E\lambda_t^{(k)}$ exists for any fixed k, and for all positive parameters. Now from (2.3), (A.2), (A.4) and (A.5) we have

$$E\left(\left|Y_t^{(k)} - Y_t^{(k-1)}\right| \mid \Delta_t = s\right) = E\left(Y_t^{(k)} - Y_t^{(k-1)} \mid \Delta_t = s\right) = E\left(\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| \mid \Delta_t = s\right)$$

and

$$\begin{aligned}
\pi_s E\left(\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| \mid \Delta_t = s\right) &= \sum_{l=1}^{r} \pi_s \left(\alpha_{sl} + \beta_{sl}\right) E\left(\left|\lambda_{t-l}^{(k-l)} - \lambda_{t-l}^{(k-l-1)}\right| \mid \Delta_t = s\right) \\
&= \sum_{l=1}^{r} \sum_{j=1}^{S} p_{js}^{(l)} \left(\alpha_{sl} + \beta_{sl}\right) \pi_j E\left(\left|\lambda_{t-l}^{(k-l)} - \lambda_{t-l}^{(k-l-1)}\right| \mid \Delta_{t-l} = j\right),
\end{aligned} \tag{A.6}$$

for all $1 \leq s \leq S$. In the case $p = q = 1$, the equality (A.6) reduces to

$$\pi_s E\left(\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| \mid \Delta_t = s\right) = \sum_{j=1}^{S} p_{js} \left(\alpha_{s1} + \beta_{s1}\right) \pi_j E\left(\left|\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)}\right| \mid \Delta_{t-1} = j\right),$$

and can then be rewritten

$$h^{(k)} = M^{(1)} h^{(k-1)},$$

where $h^{(k)} = \left(h_1^{(k)}, ..., h_S^{(k)}\right)^{\top}$ and $h_s^{(k)} = \pi_s E\left(\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| \mid \Delta_t = s\right)$. More generally, (A.6) can be embedded in a block-matrix form

$$\mathbf{h}^{(k)} = \Omega \mathbf{h}^{(k-1)}, \tag{A.7}$$

where $\mathbf{h}^{(k)} = \left(h^{(k)\top}, ..., h^{(k-r+1)\top}\right)^\top$. Under (3.3), $\mathbf{h}^{(k)} \to 0$ exponentially fast as $k \to \infty$ so $\left(\lambda_t^{(k)}\right)_k$ converges in L^1 and a.s. Moreover, since

$$\lambda_t = \psi(U_{t-1}, U_{t-2}, \ldots; \Delta_t, \Delta_{t-1}, ...),$$

where $\psi : [0,1]^\infty \times \{1, .., S\}^\infty \to [0, \infty)$ is a measurable function, the sequence (λ_t) is ergodic and so is (Y_t). $\square$

Proof of Theorem 3.2 Set $m_{s\ell} = E\left(Y_t^\ell \mid \Delta_t = s\right)$ and $\mu_{s\ell} = E\left(\lambda_t^\ell \mid \Delta_t = s\right)$ $(1 \le s \le S)$ when the moments exist, and $b_s(\ell) = \sum_{i=0}^{\ell-1} a_{s\ell}(i) E\left(\lambda_t^i \mid \Delta_t = s\right)$. Then (3.9) entails

$$\begin{aligned} m_{s\ell} &= E\left(E\left(Y_t^\ell \mid \mathcal{F}_{t-1}^a\right) \mid \Delta_t = s\right) \\ &= a_s(\ell) E\left(\lambda_t^\ell \mid \Delta_t = s\right) + b_s(\ell). \end{aligned} \tag{A.8}$$

Let us show that $EY_t^2 < \infty$ iff (3.10) holds with $\ell = 2$. For all $1 \le s \le S$ we have

$$\begin{aligned} \pi_s \mu_{s2} &= \pi_s E\left(\left(\omega_{\Delta_t} + \alpha_{\Delta_t} Y_{t-1} + \beta_{\Delta_t} \lambda_{t-1}\right)^2 \mid \Delta_t = s\right) \\ &= \pi_s \alpha_s^2 E\left(Y_{t-1}^2 \mid \Delta_t = s\right) + 2\pi_s \alpha_s \beta_s E\left(Y_{t-1}\lambda_{t-1} \mid \Delta_t = s\right) \\ &\quad + \pi_s \beta_s^2 E\left(\lambda_{t-1}^2 \mid \Delta_t = s\right) + K_s \end{aligned} \tag{A.9}$$

where

$$K_s = 2\pi_s \alpha_s \omega_s E\left(Y_{t-1} \mid \Delta_t = s\right) + 2\pi_s \beta_s \omega_s E\left(\lambda_{t-1} \mid \Delta_t = s\right) + \pi_s \omega_s^2.$$

On the other hand, by (A.8), **A1** and Lemma 3 of Francq and Zakoian (2005), $E\left(Y_{t-1}^2 \mid \Delta_t = s\right)$ in (A.9) can be rewritten as follows

$$\begin{aligned} \pi_s E\left(Y_{t-1}^2 \mid \Delta_t = s\right) &= \sum_{\tau=1}^S p_{\tau s} \pi_\tau E\left(Y_{t-1}^2 \mid \Delta_{t-1} = \tau\right) = \sum_{\tau=1}^S \pi_\tau p_{\tau s} m_{\tau 2} \\ &= \sum_{\tau=1}^S \pi_\tau p_{\tau s} \left(a_\tau(2) \mu_{\tau 2} + b_\tau(2)\right). \end{aligned} \tag{A.10}$$

Similarly,

$$\pi_s E\left(Y_{t-1}\lambda_{t-1} \mid \Delta_t = s\right) = \sum_{\tau=1}^S p_{\tau s} \pi_\tau E\left(Y_{t-1}\lambda_{t-1} \mid \Delta_{t-1} = \tau\right) = \sum_{\tau=1}^S p_{\tau s} \pi_\tau \mu_{\tau 2}$$

and

$$\pi_s E\left(\lambda_{t-1}^2 \mid \Delta_t = s\right) = \sum_{\tau=1}^{S} p_{\tau s}\pi_\tau \mu_{\tau 2}.$$

Substituting (A.10) and the two next equations into (A.9), we get

$$\pi_s \mu_{s2} = \sum_{\tau=1}^{S} p_{\tau s}\left(a_s(2)\alpha_s^2 + \beta_s^2 + 2\alpha_s\beta_s\right)\pi_\tau\mu_{\tau 2} + K_s'$$

where

$$K_s' = K_s + \pi_s\alpha_s^2 b_s(2) > 0.$$

Letting $\boldsymbol{\mu}_2 = \left(\pi_1\mu_{12}, \pi_2\mu_{22}, ..., \pi_S\mu_{S2}\right)^\top$ and $\mathbf{K}' = \left(K_1', K_2', ..., K_S'\right)^\top$, the latter equality can be embedded into the following equation

$$\boldsymbol{\mu}_2 = M_2\boldsymbol{\mu}_2 + \mathbf{K}'$$

with $\mathbf{K}' > 0$ component-wise. Therefore, in view of Lemma A.1, $EY_t^2 < \infty$ entails (3.10) for $\ell = 2$.

We now show that (3.10) is also sufficient. From Theorem 3.1 we know that

$$Y_t = \lim_{k\to\infty} \uparrow Y_t^{(k)}.$$

To prove that m_{s2} exists, it thus suffices by the monotone convergence theorem to prove that $\lim_{k\to\infty} m_{s2}^{(k)}$ is finite for all s, where $m_{su}^{(k)}$ denotes $E\left(Y_t^{(k)u} \mid \Delta_t = s\right)$ which is finite for all $u \geq 0$ and k. Letting $\mu_{su}^{(k)} = E\left(\lambda_t^{(k)u} \mid \Delta_t = s\right)$, and $b_s^{(k)}(\ell) = \sum_{i=0}^{\ell-1} a_{s\ell}(i)E\left(\lambda_t^{(k)i} \mid \Delta_t = s\right)$ we have

$$\begin{aligned}
\pi_s\mu_{s2}^{(k)} &= a_s(2)m_{s2}^{(k)} + b_s^{(k)}(2) \\
&= \sum_{\tau=1}^{S} p_{\tau s}\left(a_\tau(2)\alpha_s^2 + \left(2\alpha_s\beta_s + \beta_s^2\right)\right)\pi_\tau\mu_{\tau 2}^{(k-1)} + K_s^{(k)\prime}
\end{aligned}$$

where

$$\begin{aligned}
K_s^{(k)\prime} &= 2\pi_s\alpha_s\omega_s E\left(Y_{t-1}^{(k-1)} \mid \Delta_t = s\right) + 2\pi_s\beta_s\omega_s E\left(\lambda_{t-1}^{(k-1)} \mid \Delta_t = s\right) + \pi_s\omega_s^2 + \pi_s\alpha_s^2 b_s^{(k)}(2) \\
&\to K_s' \text{ a.s.}
\end{aligned}$$

as $k \to \infty$ because it can be shown from Theorem 3.1 that $E(Y_{t-1} \mid \Delta_t = s) = \lim_{k\to\infty} E(Y_{t-1}^{(k)} \mid \Delta_t = s)$ and $E(\lambda_{t-1} \mid \Delta_t = s) = \lim_{k\to\infty} E(\lambda_{t-1}^{(k)} \mid \Delta_t = s)$. We thus have $K_s^{(k)\prime} < cK_s'$ for all $k \geq 0$ and some $c > 0$. Letting $\boldsymbol{\mu}_2^{(k)} = \left(\pi_1 \mu_{12}^{(k)}, \pi_2 \mu_{22}^{(k)}, ..., \pi_S \mu_{S2}^{(k)}\right)^\top$ and $K^{(k)\prime} = (K_1^{(k)\prime}, K_2^{(k)\prime}, ..., K_S^{(k)\prime})^\top$, we get

$$\boldsymbol{\mu}_2^{(k)} = M_2 \boldsymbol{\mu}_2^{(k-1)} + K^{(k)\prime} \leq cK_s' \sum_{i=0}^{\infty} M_2^i.$$

It follows that $\mu_{s2} = \lim_{k\to\infty} \uparrow \mu_{s2}^{(k)} < \infty$ and then $m_{s2} = \lim_{k\to\infty} \uparrow m_{s2}^{(k)} < \infty$ under (3.10). The proof of Theorem 3.2 is complete in the case $\ell = 2$. The general case is shown by induction as for $\ell = 2$. $\square$

Proof of Theorem 3.3 Using the argument of Theorem 3.1 define $Y_t^{(k)}$ as in (A.2). Let $\lambda_t^{(k)} = 0$ if $k \leq 0$ and for $k > 0$

$$\lambda_t^{(k)} = g_{\Delta_t}(Y_{t-1}^{(k-1)}, \ldots, Y_{t-q}^{(k-q)}, \lambda_{t-1}^{(k-1)}, \ldots, \lambda_{t-p}^{(k-p)}).$$

Similarly to the proof of Theorem 3.1, to show the existence of a stationary solution it suffices to show the almost sure convergence (A.3) of $\lambda_t^{(k)}$ as $k \to \infty$. In view of (2.3) we have

$$E\left(\left|Y_t^{(k)} - Y_t^{(k-1)}\right| \mid \lambda_t^{(k)}, \lambda_t^{(k-1)}, \Delta_t = s\right) = E\left(\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| \mid \Delta_t = s\right).$$

Therefore

$$E\left(\left|Y_t^{(k)} - Y_t^{(k-1)}\right| \mid \Delta_t = s\right) = E\left(\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| \mid \Delta_t = s\right). \tag{A.11}$$

In view of (2.8), (3.15) and (A.11), we have

$$\begin{aligned}\pi_s E\left(\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| \mid \Delta_t = s\right) &\leq \sum_{l=1}^{r} \pi_s \left(\alpha_{sl} + \beta_{sl}\right) E\left(\left|\lambda_{t-l}^{(k-l)} - \lambda_{t-l}^{(k-l-1)}\right| \mid \Delta_t = s\right) \\ &= \sum_{l=1}^{r} \sum_{j=1}^{S} p_{js}^{(l)} \left(\alpha_{sl} + \beta_{sl}\right) \pi_j E\left(\left|\lambda_{t-l}^{(k-l)} - \lambda_{t-l}^{(k-l-1)}\right| \mid \Delta_{t-l} = j\right).\end{aligned} \tag{A.12}$$

Inequality (A.12) can be embedded into the following matrix inequality (element-wise)

$$\mathbf{h}^{(k)} \leq \Omega \mathbf{h}^{(k-1)},$$

where $\mathbf{h}^{(k)}$ is defined as in (A.7). Thus, under $\rho(\Omega) < 1$, $\mathbf{h}^{(k)} \to 0$ as $k \to \infty$ and $\left(\lambda_t^{(k)}\right)_k$ converges in L^1 and a.s. $\square$

Proof of Theorem 4.1 If there exists a solution to (3.2) such that $EY_t < \infty$ for all t and λ_{st} satisfies (2.9) then for some $s, \tau \in \{1, .., S\}$,

$$E(\lambda_{st} \mid \Delta_t = \tau) = \omega_s + \sum_{i=1}^{q} \alpha_{si} E(Y_{t-i} \mid \Delta_t = \tau) + \sum_{j=1}^{p} \beta_{sj} E(\lambda_{s,t-j} \mid \Delta_t = \tau). \tag{A.13}$$

Under **A1**, Lemma 3 of Francq and Zakoian (2005) entails

$$\begin{aligned} E(\lambda_{s,t-j} \mid \Delta_t = \tau) &= \sum_{s'=1}^{S} P(\Delta_{t-j} = s' \mid \Delta_t = \tau) E(\lambda_{s,t-j} \mid \Delta_t = \tau, \Delta_{t-j} = s') \\ &= \sum_{s'=1}^{S} p_{\tau s'}^{(-j)} E(\lambda_{s,t-j} \mid \Delta_{t-j} = s') \\ &= \sum_{s'=1}^{S} \tfrac{\pi_{s'}}{\pi_\tau} p_{s'\tau}^{(j)} E(\lambda_{s,t-j} \mid \Delta_{t-j} = s'), \quad j = 1, ..., p, \end{aligned} \tag{A.14}$$

and

$$\begin{aligned} E(Y_{t-i} \mid \Delta_t = \tau) &= \sum_{s'=1}^{S} P(\Delta_{t-i} = s' \mid \Delta_t = \tau) E(Y_{t-i} \mid \Delta_{t-i} = s'), \\ &= \sum_{s'=1}^{S} \tfrac{\pi_{s'}}{\pi_\tau} p_{s'\tau}^{(i)} E(\lambda_{s',t-i} \mid \Delta_{t-i} = s'), \quad i = 1, ..., q. \end{aligned} \tag{A.15}$$

Thus equality (A.13) can be written as follows

$$E(\lambda_{st} \mid \Delta_t = \tau) = \omega_s + \sum_{l=1}^{r} \sum_{s'=1}^{S} \tfrac{\pi_{s'}}{\pi_\tau} p_{s'\tau}^{(l)} \left[\alpha_{sl} E(\lambda_{s',t-l} \mid \Delta_{t-l} = s') + \beta_{sl} E(\lambda_{s,t-l} \mid \Delta_{t-l} = s')\right]. \tag{A.16}$$

Letting $w_{s\tau} = \pi_\tau E(\lambda_{st} \mid \Delta_t = \tau)$, $w = (w_{11}, w_{21}, ..., w_{SS})^\top$, $\mathbf{w} = \left(w^\top, ..., w^\top\right)^\top_{rS^2 \times 1}$, $\omega = (\omega_1, ..., \omega_S)^\top$, $\varpi = \left(\pi_1 \omega^\top, ..., \pi_S \omega^\top\right)^\top_{S^2 \times 1}$, and $\boldsymbol{\varpi} = (\varpi^\top, 0^\top_{S^2 \times 1}, ..., 0^\top_{S^2 \times 1})^\top$, (A.16) can be cast in the following block-matrix form

$$\mathbf{w} = \boldsymbol{\varpi} + D\mathbf{w}.$$

Therefore, in view of Lemma A.1, condition (4.2) holds.

To show the sufficiency of (4.2), let (U_t) be defined as above and set $Y_t^{(k)} = \lambda_t^{(k)} = 0$ when $k \leq 0$, and for $k > 0$

$$Y_t^{(k)} = \sum_{s=1}^{S} 1_{[\Delta_t = s]} F^{-}_{s,\lambda_{st}^{(k)}}(U_t), \qquad \lambda_{st}^{(k)} = \omega_s + \sum_{i=1}^{q} \alpha_{si} Y_{t-i}^{(k-i)} + \sum_{j=1}^{p} \beta_{sj} \lambda_{s,t-j}^{(k-j)}. \tag{A.17}$$

It can be shown by induction using the above arguments that conditionally on $\Delta_t = s$,

$$0 \leq \lambda_{st}^{(k-1)} \leq \lambda_{st}^{(k)} \quad 0 \leq Y_t^{(k-1)} = F^{-}_{s,\lambda_{st}^{(k-1)}}(U_t) \leq Y_t^{(k)}, \quad 1 \leq s \leq S$$

so the sequences $(\lambda_{st}^{(k)})_k$ and $(Y_t^{(k)})_k$ are non-decreasing for all $1 \leq s \leq S$. Moreover, similarly to (A.14) and (A.15), it can be seen that

$$E\left(\left|Y_{t-i}^{(k-i)} - Y_{t-i}^{(k-i-1)}\right| \mid \Delta_t = \tau\right) = \sum_{s'=1}^{S} p_{\tau s'}^{(-i)} E\left(\left|\lambda_{s',t-i}^{(k-i)} - \lambda_{s',t-j}^{(k-j-1)}\right| \mid \Delta_{t-i} = s'\right),\ i = 1, ..., q,$$

$$E\left(\left|\lambda_{s,t-j}^{(k-j)} - \lambda_{s,t-j}^{(k-j-1)}\right| \mid \Delta_t = \tau\right) = \sum_{s'=1}^{S} p_{\tau s'}^{(-j)} E\left(\left|\lambda_{s,t-j}^{(k-j)} - \lambda_{s,t-j}^{(k-j-1)}\right| \mid \Delta_{t-j} = s'\right),\ j = 1, ..., p.$$

Therefore

$$\pi_\tau E\left(\left|\lambda_{st}^{(k)} - \lambda_{st}^{(k-1)}\right| \mid \Delta_t = \tau\right) =$$
$$\sum_{l=1}^{r} \sum_{s'=1}^{S} \pi_{s'} p_{s'\tau}^{(l)} \left[\alpha_{sl} E\left(\left|\lambda_{s',t-l}^{(k-l)} - \lambda_{s',t-l}^{(k-l-1)}\right| \mid \Delta_{t-l} = s'\right) + \beta_{sl} E\left(\left|\lambda_{s,t-l}^{(k-l)} - \lambda_{s,t-l}^{(k-l-1)}\right| \mid \Delta_{t-l} = s'\right)\right]. \tag{A.18}$$

Setting $w_{s\tau}^{(k)} = \pi_\tau E\left(\left|\lambda_{st}^{(k)} - \lambda_{st}^{(k-1)}\right| \mid \Delta_t = \tau\right)$, $w^{(k)} = \left(w_{11}^{(k)}, w_{21}^{(k)}, ..., w_{SS}^{(k)}\right)^\top$, and $\mathbf{w}^{(k)} = \left(w^{(k)\top}, ..., w^{(k-r+1)\top}\right)^\top$, (A.18) can be stacked in the following linear system

$$\mathbf{w}^{(k)} = D\mathbf{w}^{(k)}.$$

It follows under (4.2) that $\mathbf{w}^{(k)} \to 0$ exponentially fast as $k \to \infty$, so $(Y_t^{(k)})_k$ is a Cauchy sequence in L^1 and thus converges in L^1 and a.s. □

Proof of Theorem 4.3 Define $Y_t^{(k)}$ as in (A.17). Let $\lambda_{st}^{(k)} = 0$ if $k \leq 0$ and when $k > 0$

$$\lambda_{st}^{(k)} = g_s(Y_{t-1}^{(k-1)}, \ldots, Y_{t-q}^{(k-q)}, \lambda_{s,t-1}^{(k-1)}, \ldots, \lambda_{s,t-p}^{(k-p)}), \quad 1 \leq s \leq S.$$

By the same arguments used in (A.14), (A.15) and (A.18), it follows that

$$E\left(\left|\lambda_{st}^{(k)}-\lambda_{st}^{(k-1)}\right| \mid \Delta_t=\tau\right) \leq$$
$$\sum_{l=1}^{r}\sum_{s'=1}^{S}\tfrac{\pi_{s'}}{\pi_\tau}p_{s'\tau}^{(l)}\left[\alpha_{sl}E\left(\left|\lambda_{s',t-l}^{(k-1)}-\lambda_{s',t-l}^{(k-2)}\right| \mid \Delta_{t-l}=s'\right)+\beta_{sl}E\left(\left|\lambda_{s,t-l}^{(k-1)}-\lambda_{s,t-l}^{(k-2)}\right| \mid \Delta_{t-l}=s'\right)\right].$$

The latter inequality can be embedded as follows

$$\mathbf{w}^{(k)} \leq D\mathbf{w}^{(k)},$$

where $\mathbf{w}^{(k)}$ is defined in (A.18). Thus the conclusion follows under (4.2). $\square$

Proof of Theorem 5.1 Let (Y_t) be a solution of (2.6) with λ_{st} given by (2.12) such that $EY_t<\infty$ for all t. For some $s\in\{1,..,S\}$ we have

$$\begin{aligned}
&E\left(\lambda_{st} \mid \Delta_t=s\right)\\
&=\omega_s+\sum_{i=1}^{q}\alpha_{si}E\left(Y_{t-i} \mid \Delta_t=s\right)+\sum_{j=1}^{p}\beta_{sj}E\left(E\left(\lambda_{\Delta_{t-j},t-j} \mid \Delta_t=s,\mathcal{F}_{t-1}\right) \mid \Delta_t=s\right).\\
&=\omega_s+\sum_{i=1}^{q}\alpha_{si}E\left(Y_{t-i} \mid \Delta_t=s\right)+\sum_{j=1}^{p}\beta_{sj}E\left(\lambda_{\Delta_{t-j},t-j} \mid \Delta_t=s\right).
\end{aligned}\tag{A.19}$$

Under **A1**, using Lemma 3 in Francq and Zakoian (2005), it follows that

$$E\left(\lambda_{\Delta_{t-j},t-j} \mid \Delta_t=s\right)=\sum_{\tau=1}^{S}P\left(\Delta_{t-j}=\tau \mid \Delta_t=s\right)E\left(\lambda_{\tau,t-j} \mid \Delta_{t-j}=\tau\right).\tag{A.20}$$

and

$$\begin{aligned}
E\left(Y_{t-i} \mid \Delta_t=s\right)&=\sum_{\tau=1}^{S}P\left(\Delta_{t-i}=\tau \mid \Delta_t=s\right)E\left(Y_{t-i} \mid \Delta_{t-i}=\tau\right)\\
&=\sum_{\tau=1}^{S}p_{s\tau}^{(-i)}\sum_{s'=1}^{S}P\left(\Delta_{t-i}=s' \mid \Delta_{t-i}=\tau\right)E\left(\lambda_{s',t-i} \mid \Delta_{t-i}=\tau\right)\\
&=\sum_{\tau=1}^{S}p_{s\tau}^{(-i)}E\left(\lambda_{\tau,t-i} \mid \Delta_{t-i}=\tau\right).
\end{aligned}\tag{A.21}$$

Substituting (A.20) and (A.21) into (A.19), we thus obtain

$$\begin{aligned}
E\left(\lambda_{st} \mid \Delta_t=s\right)&=\omega_s+\sum_{l=1}^{r}\sum_{\tau=1}^{S}p_{s\tau}^{(-l)}\left(\alpha_{sl}+\beta_{sl}\right)E\left(\lambda_{\tau,t-l} \mid \Delta_{t-l}=\tau\right)\\
&=\omega_s+\sum_{l=1}^{r}\sum_{\tau=1}^{S}\tfrac{\pi_\tau}{\pi_s}p_{\tau s}^{(l)}\left(\alpha_{sl}+\beta_{sl}\right)E\left(\lambda_{\tau,t-l} \mid \Delta_{t-l}=\tau\right),
\end{aligned}$$

which can be embedded into the following block-matrix form

$$\mathbf{v} = \mathbf{d} + \Omega \mathbf{v},$$

where $\mathbf{v} = \left(v^\top, ..., v^\top\right)^\top_{rS\times 1}$, $v = (v_1, ..., v_S)^\top$, $v_s = \pi_s E\left(\lambda_{st} \mid \Delta_t = s\right)$, and $\mathbf{d}$ is defined in (A.1). Therefore, Lemma A.1 and already given arguments entail (3.3).

For the sufficiency of (3.3), let $Y_t^{(k)} = F^-_{\Delta_t, \lambda^{(k)}_{\Delta_t,t}}(U_t)$ and $\mathcal{F}_t^{(k)}$ be the sigma-field generated by $\{Y_{t-u}^{(k-u)}, u \geq 0\}$. Take $\lambda_t^{(k)} = 0$ when $k \leq 0$ and

$$\lambda_{st}^{(k)} = \omega_s + \sum_{i=1}^q \alpha_{si} Y_{t-i}^{(k-i)} + \sum_{j=1}^p \beta_{sj} E\left(\lambda^{(k-j)}_{\Delta_{t-j},t-j} \mid \Delta_t = s, \mathcal{F}_{t-1}^{(k-1)}\right), \tag{A.22}$$

when $k > 0$. To compute the conditional expectation, we note that, by already given arguments,

$$E\left(\lambda^{(k-j)}_{\Delta_{t-j},t-j} \mid \Delta_t = s, \mathcal{F}_{t-1}^{(k-1)}\right) = \sum_{\tau=1}^S \lambda^{(k-j)}_{\tau,t-j} P(\Delta_{t-j} = \tau \mid \Delta_t = s, \mathcal{F}_{t-1}^{(k-1)})$$

and we adapt the algorithm of Hamilton (1994, p. 692-694) to compute recursively the conditional probabilities. Using the above arguments, it is easy to show by induction that the sequences $(Y_t^{(k)})_k$ and $(\lambda_{st}^{(k)})_k$ are nondecreasing for all $1 \leq s \leq S$. Moreover, in view of (A.22) we have

$$\begin{aligned} E\left(\left|\lambda_{st}^{(k)} - \lambda_{st}^{(k-1)}\right| \mid \Delta_t = s\right) &= \sum_{i=1}^q \alpha_{si} E\left(\left|Y_{t-i}^{(k-i)} - Y_{t-i}^{(k-i-1)}\right| \mid \Delta_t = s\right) \\ &+ \sum_{j=1}^p \beta_{sj} E\left(\left|\lambda^{(k-j)}_{\Delta_{t-j},t-j} - \lambda^{(k-j-1)}_{\Delta_{t-j},t-j}\right| \mid \Delta_t = s, \mathcal{F}_{t-1}\right). \end{aligned} \tag{A.23}$$

Similarly to (A.20)-(A.21), the expectations in the right hand side of (A.23) can be rewritten as

$$E\left(\left|\lambda^{(k-j)}_{\Delta_{t-j},t-j} - \lambda^{(k-j-1)}_{\Delta_{t-j},t-j}\right| \mid \Delta_t = s, \mathcal{F}_{t-1}\right) = \sum_{\tau=1}^S p_{s\tau}^{(-j)} E\left(\left|\lambda^{(k-j)}_{\tau,t-j} - \lambda^{(k-j-1)}_{\tau,t-j}\right| \mid \Delta_{t-j} = \tau\right) \tag{A.24}$$

and

$$E\left(\left|Y_{t-i}^{(k-i)} - Y_{t-i}^{(k-i-1)}\right| \mid \Delta_t = s\right) = \sum_{\tau=1}^S p_{s\tau}^{(-i)} E\left(\left|\lambda^{(k-i)}_{\tau,t-i} - \lambda^{(k-i-1)}_{\tau,t-i}\right| \mid \Delta_{t-i} = \tau\right). \tag{A.25}$$

Therefore, substituting (A.24) and (A.25) into (A.23) we get

$$E\left(\left|\lambda_{st}^{(k)}-\lambda_{st}^{(k-1)}\right| \mid \Delta_t=s\right)=\sum_{l=1}^{r}\sum_{\tau=1}^{S}\tfrac{\pi_\tau}{\pi_s}p_{\tau s}^{(l)}\left(\alpha_{sl}+\beta_{sl}\right)E\left(\left|\lambda_{\tau,t-l}^{(k-l)}-\lambda_{\tau,t-l}^{(k-l-1)}\right| \mid \Delta_{t-l}=\tau\right),$$

which can be embedded into the following system

$$\mathbf{v}^{(k)}=\Omega\mathbf{v}^{(k-1)}, \tag{A.26}$$

where $\mathbf{v}^{(k)}=\left(v^{(k)\top},...,v^{(k)\top}\right)^\top_{rS\times1}$, $v^{(k)}=\left(v_1^{(k)},...,v_S^{(k)}\right)^\top$, and $v_s^{(k)}=\pi_s E\left(\left|\lambda_{st}^{(k)}-\lambda_{st}^{(k-1)}\right| \mid \Delta_t=s\right)$. Thus under (3.3), $\mathbf{v}^{(k)}\to 0$ exponentially fast as $k\to\infty$ so $(Y_t^{(k)})_k$ converges in L^1 and a.s.

□

Proof of Theorem 5.2 Let $Y_t^{(k)}$ be defined as in (A.17) and $\lambda_t^{(k)}$ given by

$$\lambda_{st}^{(k)}=\begin{cases} g_s(Y_{t-1}^{(k-1)},\dots,Y_{t-q}^{(k-q)},\mu_{t-1}^{(k-1)},\dots,\mu_{t-p}^{(k-p)}) \text{ if } k>0 \\ 0 \qquad \text{if } k\le 0, \end{cases}$$

where $\mu_{t-j}^{(k-j)}=E\left(\lambda_{\Delta_{t-j},t-j}^{(k-j)} \mid \Delta_t=s,\mathcal{F}_{t-1}\right)$. It can be easily shown by induction that $\left(Y_t^{(k)}\right)_k$ and $\left(\lambda_{st}^{(k)}\right)_k$ are nondecreasing. Moreover, similarly to (A.20)-(A.21) and (A.24)-(A.25) we can write

$$E\left(\left|\lambda_{\Delta_{t-j},t-j}^{(k-j)}-\lambda_{\Delta_{t-j},t-j}^{(k-j-1)}\right| \mid \Delta_t=s,\mathcal{F}_{t-1}\right)\le\sum_{\tau=1}^{S}p_{s\tau}^{(-j)}E\left(\left|\lambda_{\tau,t-j}^{(k-j)}-\lambda_{\tau,t-j}^{(k-j-1)}\right| \mid \Delta_{t-j}=\tau\right),$$

and

$$E\left(\left|Y_{t-i}^{(k-i)}-Y_{t-i}^{(k-i-1)}\right| \mid \Delta_t=s\right)\le\sum_{\tau=1}^{S}p_{s\tau}^{(-i)}E\left(\left|\lambda_{\tau,t-i}^{(k-i)}-\lambda_{\tau,t-i}^{(k-i-1)}\right| \mid \Delta_{t-i}=\tau\right).$$

Therefore, by the Lipschitz property (3.15) of the functions $(g_s)_s$, it follows that

$$E\left(\left|\lambda_{st}^{(k)}-\lambda_{st}^{(k-1)}\right| \mid \Delta_t=s\right)\le\sum_{l=1}^{r}\sum_{\tau=1}^{S}\tfrac{\pi_\tau}{\pi_s}p_{\tau s}^{(l)}\left(\alpha_{sl}+\beta_{sl}\right)E\left(\left|\lambda_{\tau,t-l}^{(k-l)}-\lambda_{\tau,t-l}^{(k-l-1)}\right| \mid \Delta_{t-l}=\tau\right),$$

which can be cast in the following system of inequalities

$$\mathbf{v}^{(k)}\le\Omega\mathbf{v}^{(k-1)},$$

where $\mathbf{v}^{(k)}$ is defined as in (A.26). Thus under (3.3) it follows that $\mathbf{v}^{(k)}\to 0$ exponentially fast as $k\to\infty$ so $(Y_t^{(k)})_k$ converges in L^1 and a.s. □

References

[1] Abramson, A. and Cohen, I. (2007). On the stationarity of Markov-switching GARCH processes. *Econometric Theory* 23, 485–500.

[2] Aknouche, A. and Demmouche, N. (2019). Ergodicity conditions for a double mixed Poisson autoregression. *Statistics and Probability Letters*, 147, 6–11.

[3] Aknouche, A. and Francq, C. (2020). Count and duration time series with equal conditional stochastic and mean orders. *Econometric Theory*, DOI: https://doi.org/10.1017/S0266466620000134.

[4] Bauwens, L., Preminger, A., and Rombouts, J.V.K. (2010). Theory and inference for a Markov switching GARCH model. *Econometrics Journal* 13, 218–244.

[5] Benjamin, M.A., Rigby, R.A. and Stasinopoulos, D.M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association* 98, 214–223.

[6] Berentsen, G.D., Bulla, J., Maruotti, A. and Støve, B. (2018). Modelling corporate defaults: A Markov-switching Poisson log-linear autoregressive model. *ArXiv*, https://arxiv.org/abs/1804.09252.

[7] Bhogal, S.K. and Variyam, R.T. (2019). Conditional duration models for high-frequency data: A review on recent developments. *Journal of Economic Surveys* 33, 252–273.

[8] Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization* 31, 942–959.

[9] Chen, F., Diebold, F.X. and Schorfheide, F. (2013). A Markov-switching multifractal inter-trade duration model, with application to US equities. *Journal of Econometrics* 177, 320–342.

[10] Creal, D., Koopman, S.J. and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28, 777–795.

[11] Davis, R.A. and Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica* 26, 1673–1707.

[12] De Luca, G. and Zuccolotto, P. (2006). Regime-switching Pareto distributions for ACD models. *Computational Statistics & Data Analysis* 51, 2179–2191.

[13] Diop, M.L., Diop, A. and Diongue, A.K. (2016). A mixture integer-valued GARCH model. *Revstat - Statistical Journal* 14, 245–271.

[14] Diop, M.L., Diop, A. and Diongue, A.K. (2018). A negative binomial mixture integer-valued GARCH model. *Statistika Afrika* 13, 1645–1666.

[15] Doukhan, P., Fokianos, K. and Rynkiewicz, J. (2018). Mixtures of Nonlinear Poisson Autoregressions. Working paper, ResearchGate 326478827.

[16] Doukhan, P. and Wintenberger, O. (2008). Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications* 118, 1997–2013.

[17] Engle, R. (2002). New frontiers for Arch models. *Journal of Applied Econometrics* 17, 425–446.

[18] Engle, R. and Russell, J. (1998). Autoregressive conditional duration: A new model for irregular spaced transaction data. *Econometrica* 66, 1127–1162.

[19] Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis* 27, 923–942.

[20] Fisher, R.A. (1953). Dispersion on a Sphere. *Proceedings of the Royal Society of London A* 217, 295–305.

[21] Fong, W.M. and See, K.H. (2001). Modelling the conditional volatility of commodity index futures as a regime switching process. *Journal of Applied Econometrics* 16, 133–163.

[22] Francq, C. and Roussignol, M. (1998). Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum-likelihood estimator, *Statistics* 32, 151–173.

[23] Francq, C., and Thieu, L. (2019). Qml inference for volatility models with covariates. *Econometric Theory* 35, 37–72.

[24] Francq, C. and Zakoian, J.-M. (2001). Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics* 102, 339–364.

[25] Francq, C. and Zakoian, J.-M. (2005). The L2-structures of standard and switching-regime GARCH models. *Stochastic Processes and their Applications* 115, 1557–1582.

[26] Francq, C. and Zakoian, J.-M. (2008). Deriving the autocovariances of powers of Markov-switching GARCH models, with applications to statistical inference. *Computational Statistics and Data Analysis* 52, 3027–3046.

[27] Francq, C. and Zakoian, J.-M. (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*, 2nd edition, John Wiley & Sons.

[28] Fokianos, K., Rahbek, A., and Tjøstheim D. (2009). Poisson autoregression. *Journal of the American Statistical Association* 140, 1430–1439.

[29] Gonçalves, E., Mendes-Lopes, N. and Silva, F. (2015). Infinitely divisible distributions in integer-valued GARCH models. *Journal of Time Series Analysis* 36, 503–527.

[30] Gorgi, P. and Koopman, S.J. (2020). Beta observation-driven models with exogenous regressors: A joint analysis of realized correlation and leverage effects. *Tinbergen Institute Discussion Paper 20-004/III.*

[31] Gray, S. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42, 27–62.

[32] Haas, M., Mittnik, S. and Paolella, M. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics* 2, 493–530.

[33] Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.

[34] Hamilton, J.D. (1994). Time Series Analysis. Princeton, New Jersey: Princeton University Press.

[35] Hamilton, J.D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64, 307–333.

[36] Harvey, A.C. (2013). Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series. Cambridge University Press.

[37] Hautsch, N. (2012). *Econometrics of financial high-frequency data*. Springer-Verlag.

[38] Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. *Available at SSRN 1117187.*

[39] Horn, R. and Johnson, C.R. (2013). *Matrix analysis*, 2nd edition, Cambridge University Press.

[40] Hujer, R. and Vuletic, S. (2007). Econometric analysis of financial trade processes by discrete mixture duration models. *Journal of Economic Dynamics and Control* 31, 635–667.

[41] Hujer, R., Vuletic, S. and Kokot, S. (2002). The Markov Switching ACD Model. Finance and Accounting Working Paper 90, Johann Wofgang Goethe-University, Frankfurt.

[42] Hujer, R., Kokot, S. and Vuletic, S. (2003). Comparison of MSACD models. Johann Wolfgang Goethe University, Frankfurt/Main.

[43] Jorgensen, B. (1997). *The theory of dispersion models*. Chapman and Hall, London.

[44] Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics* 27, 363–94.

[45] Lee, J. and Hwang, E. (2018). A generalized regime-switching integer-valued GARCH(1, 1) model and its volatility forecasting. *Communications for Statistical Applications and Methods* 25, 29–42.

[46] Liu, J.-C. (2006). Stationarity of a Markov-switching GARCH model. *Journal of Financial Econometrics* 4, 573–593.

[47] Mao, H., Zhu, F. and Cui, Y. (2019). Mixtures of Nonlinear Poisson Autoregressions. *Statistical Methods & Applications*, forthcoming. https://doi.org/10.1007/s10260-019-00498-2.

[48] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models* (2nd edn). London: Chapman and Hall.

[49] Nelder, J.A. and Wedderburn, R.W. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society, Series A* 135, 370–384.

[50] Rydberg, T.H. and Shephard, N. (2000). BIN models for trade-by-trade data. Modelling the number of trades in a fixed interval of time. In World Conference Econometric Society, 2000, Seattle. *Contributed Paper 0740.*

[51] Zheng, T., Xiao, H. and Chen, R. (2015). Generalized ARMA with martingale difference errors. *Journal of Econometrics* 189, 492–506.

[52] Zhu, F., Li, Q. and Wang, D. (2010). A mixture integer-valued ARCH model. *Journal of Statistical Planning and Inference* 140, 2025–2036.

Supplement to "Stationarity and ergodicity of Markov switching positive conditional mean models"

Abdelhakim Aknouche and Christian Francq

Proof of Proposition 3.1

If there exists a stationary solution of (3.2) such that $m = EY_t = E\lambda_t > 0$ then

$$\left(1 - \sum_{s=1}^{S} \pi_s \left(\sum_{i=1}^{q} \alpha_{si} + \sum_{j=1}^{p} \beta_{sj}\right)\right) m = \sum_{s=1}^{S} \pi_s \omega_s.$$

Therefore, by the positivity of the parameters, (3.6) should be satisfied.

Conversely, assume (3.6) holds and let (U_t) be an iid sequence of random variables uniformly distributed in $[0,1]$, independent of the sequence (Δ_t). Let also $Y_t^{(k)}$ and $\lambda_t^{(k)}$ $(k \in \mathbb{N})$ be defined as in (A.2). To show the existence of a solution of (3.2) it is sufficient to show that

$$\lambda_t = \lim_{k\to\infty} \lambda_t^{(k)} \text{ exists almost surely (a.s.) in } [0, +\infty).$$

By the same argument used in proving (A.4) and (A.5) we have for all k

$$0 \leq \lambda_t^{(k-1)} \leq \lambda_t^{(k)} \quad \text{a.s.}$$

and

$$E\left(Y_t^{(k)} - Y_t^{(k-1)}\right) = E\left(\lambda_t^{(k)} - \lambda_t^{(k-1)}\right) \in [0, \infty).$$

Therefore, it follows under (3.6) that

$$E\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| = \sum_{s=1}^{S} \pi_s \sum_{i=1}^{r} (\alpha_{si} + \beta_{si}) E\left|\lambda_{t-i}^{(k-i)} - \lambda_{t-i}^{(k-i-1)}\right| \leq K\rho^k, \quad \forall k \geq 1,$$

where $K > 0$ and $\rho \in (0,1)$. This implies that $\left(\lambda_t^{(k)}\right)_k$ converges in L^1 and a.s., establishing the result. $\square$

Proof of Proposition 3.2

Using the argument in Proposition 3.1 define $Y_t^{(k)}$ as in (A.2) and

$$\lambda_t^{(k)} = \begin{cases} g_{\Delta_t}(Y_{t-1}^{(k-1)}, \ldots, Y_{t-q}^{(k-q)}, \lambda_{t-1}^{(k-1)}, \ldots, \lambda_{t-p}^{(k-p)}; \theta_{\Delta_t}) & \text{if } k > 0 \\ 0 \quad \text{if } k \leq 0. \end{cases}$$

As in the proof of Proposition 3.1, to show the existence of a stationary solution it suffices to show the almost sure convergence of $\lambda_t^{(k)}$ as $k \to \infty$. In view of (2.3) we have

$$E\left(\left|Y_t^{(k)} - Y_t^{(k-1)}\right| \Big/ \lambda_t^{(k)}, \lambda_t^{(k-1)}\right) = \left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right|.$$

Therefore

$$E\left|Y_t^{(k)} - Y_t^{(k-1)}\right| = E\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right|.$$

Under (3.15) and (3.3) it follows that

$$E\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| \leq \sum_{s=1}^{S} \sum_{l=1}^{\max(p,q)} \pi_s \left(\alpha_{sl} + \beta_{sl}\right) E\left|\lambda_{t-l}^{(k-l)} - \lambda_{t-l}^{(k-l-1)}\right| \leq K\rho^k, \quad \forall k \geq 1$$

for some constant $K > 0$ and $\rho \in (0,1)$. The proof of the existence of a stationary and ergodic solution thus follows. $\square$

Proof of Theorem 4.2

If there exists a solution of (2.6) such that $EY_t < \infty$ and $E\lambda_{st} < \infty$ $(1 \leq s \leq S)$ then $EY_t = \sum_{s=1}^{S} \pi_s E\lambda_{st}$ and

$$E\lambda_{st} = \omega_s + \sum_{i=1}^{q} \alpha_{si} \sum_{\tau=1}^{S} \pi_\tau E\lambda_{\tau,t-i} + \sum_{j=1}^{p} \beta_{sj} E\lambda_{s,t-j}, \quad 1 \leq s \leq S. \tag{1}$$

Letting $v=(E\lambda_{1t},...,E\lambda_{St})^\top$, $\mathbf{v}=\left(v^\top,...,v^\top\right)^\top_{rS\times 1}$ and $\boldsymbol{\omega}=\left(\omega^\top,0^\top_{S\times 1},...,0^\top_{S\times 1}\right)^\top_{rS\times 1}$, equality (1) can be embedded in the following block-matrix form

$$\mathbf{v}_S=\boldsymbol{\omega}_S+\Sigma\mathbf{v}_S$$

where $\mathbf{v}_S=\Sigma^{S-1}\mathbf{v}>0$ and $\boldsymbol{\omega}_S=\Sigma^{S-1}\boldsymbol{\omega}>0$. Therefore, by Lemma A.1, the condition (4.7) should be satisfied.

Conversely, let (U_t) be an iid sequence of random variables uniformly distributed in $[0,1]$, independent of the sequence (λ_{st}). Let also $Y_t^{(k)}=0=\lambda_t^{(k)}$ when $k<0$, and for $k\geq 0$,

$$Y_t^{(k)}=\sum_{s=1}^{S}1_{[\Delta_t=s]}F^-_{s,\lambda_{st}^{(k)}}(U_t)\qquad \lambda_{st}^{(k)}=\omega_s+\sum_{i=1}^{q}\alpha_{si}Y_{t-i}^{(k-i)}+\sum_{j=1}^{p}\beta_{sj}\lambda_{s,t-j}^{(k-j)}. \tag{2}$$

Conditionally on $\Delta_t=s$, it can be shown by induction that for all $1\leq s\leq S$,

$$\begin{aligned}&0\leq\lambda_{st}^{(k-1)}\leq\lambda_{st}^{(k)},\\&0\leq Y_t^{(k-1)}=F^-_{s,\lambda_{st}^{(k-1)}}(U_t)\leq Y_t^{(k)},\end{aligned}$$

so the sequences $(\lambda_{st}^{(k)})_k$ and $(Y_t^{(k)})_k$ are non-decreasing for all $1\leq s\leq S$. Since

$$E\left(\left|Y_t^{(k)}-Y_t^{(k-1)}\right|\right)=E\left|\lambda_{st}^{(k)}-\lambda_{st}^{(k-1)}\right|,\ 1\leq s\leq S,$$

we then have

$$E\left|Y_t^{(k)}-Y_t^{(k-1)}\right|=E\left(E\left|Y_t^{(k)}-Y_t^{(k-1)}\right|\mid\Delta_t\right)=\sum_{s=1}^{S}\pi_sE\left|\lambda_{st}^{(k)}-\lambda_{st}^{(k-1)}\right|.$$

Therefore

$$E\left|\lambda_{st}^{(k)}-\lambda_{st}^{(k-1)}\right|=\sum_{i=1}^{q}\alpha_{si}\sum_{\tau=1}^{S}\pi_\tau E\left|\lambda_{\tau,t-i}^{(k-i)}-\lambda_{\tau,t-i}^{(k-i-1)}\right|+\sum_{j=1}^{p}\beta_{sj}E\left|\lambda_{s,t-j}^{(k-j)}-\lambda_{s,t-j}^{(k-j-1)}\right|.$$

Letting $u^{(k)}=\left(E\left|\lambda_{1t}^{(k)}-\lambda_{1t}^{(k-1)}\right|,...,E\left|\lambda_{St}^{(k)}-\lambda_{St}^{(k-1)}\right|\right)^\top$, the latter equality can be cast in the following block-matrix equality

$$\mathbf{u}^{(k)}=\Sigma\mathbf{u}^{(k-1)}, \tag{3}$$

where $\mathbf{u}^{(k)} = \left(u^{(k)\top}, ..., u^{(k-r+1)\top}\right)^\top$. Therefore, under the condition (4.7)

$$\mathbf{u}^{(k)} \to 0$$

exponentially fast as $k \to \infty$. Hence $(Y_t^{(k)})_k$ is a Cauchy sequence in L^1 and thus converges in L^1 and a.s. $\square$

Proof of Theorem 4.4

Let $Y_t^{(k)}$ be defined as in (2) and $\lambda_{st}^{(k)}$ $(1 \le s \le S)$ be given by

$$\lambda_{st}^{(k)} = \begin{cases} g_s(Y_{t-1}^{(k-1)}, \dots, Y_{t-q}^{(k-q)}, \lambda_{s,t-1}^{(k-1)}, \dots, \lambda_{s,t-p}^{(k-p)}), & \text{if } k > 0 \\ 0 \qquad \text{if } k \le 0. \end{cases}$$

Using the above arguments it can be easily shown that

$$E\left(\left|Y_t^{(k)} - Y_t^{(k-1)}\right| \mid \boldsymbol{\lambda}_t^{(k)}, \boldsymbol{\lambda}_t^{(k-1)}\right) \le \sum_{s=1}^{S} \pi_s |\lambda_{st}^{(k)} - \lambda_{st}^{(k-1)}|,$$

where $\boldsymbol{\lambda}_t^{(k)} = \left(\lambda_{1t}^{(k)}, ..., \lambda_{St}^{(k)}\right)'$. Therefore

$$E\left|Y_t^{(k)} - Y_t^{(k-1)}\right| \le \sum_{s=1}^{S} \pi_s E\left|\lambda_{st}^{(k)} - \lambda_{st}^{(k-1)}\right|$$

and

$$E\left|\lambda_{st}^{(k)} - \lambda_{st}^{(k-1)}\right| \le \sum_{i=1}^{q} \alpha_{si} \sum_{\tau=1}^{S} \pi_\tau E\left|\lambda_{\tau,t-i}^{(k-i)} - \lambda_{\tau,t-i}^{(k-i-1)}\right| + \sum_{j=1}^{p} \beta_{sj} E\left|\lambda_{s,t-j}^{(k-j)} - \lambda_{s,t-j}^{(k-j-1)}\right|.$$

Using the same notations as in (3), the latter inequality becomes

$$\mathbf{u}^{(k)} \le \Sigma \mathbf{u}^{(k-1)},$$

(element-wise) from which the conclusion follows under the condition (4.7). $\square$