On Categorical Time Series Models With Covariates

Konstantinos Fokianos[*] Lionel Truquet [†‡]

September 2017

Abstract

We study the problem of stationarity and ergodicity for autoregressive multinomial logistic time series models which possibly include a latent process and are defined by a GARCH-type recursive equation. We improve considerably upon the existing results related to stationarity and ergodicity conditions of such models. Proofs are based on theory developed for chains with complete connections. This approach is based on a useful coupling technique which is utilized for studying ergodicity of more general finite-state stochastic processes. Such processes generalize finite-state Markov chains by assuming infinite order models of past values. For finite order Markov chains, we also discuss ergodicity properties when some strongly exogenous covariates are considered in the dynamics of the process.

2010 Mathematics Subject Classification: Primary 62M10; secondary 60G10, 60B12.

Keywords and Phrases: autoregression, categorical data, chains with complete connection, coupling, covariates, ergodicity, Markov chains.

[*]University of Cyprus, Department of Mathematics & Statistics, PO BOX 20537, 1678 Nicosia, Cyprus *Email: fokianos@ucy.ac.cy*.

[†]UMR 6625 CNRS IRMAR, University of Rennes 1, Campus de Beaulieu, F-35042 Rennes Cedex, France and

[‡]ENSAI, Campus de Ker-Lann, rue Blaise Pascal, BP 37203, 35172 Bruz cedex, France. *Email: lionel.truquet@ensai.fr*.

1 Introduction

The goal of this article is to improve upon theoretical properties of regression based models for the analysis of categorical time series that might include some covariates. Binary time series are particular cases of a categorical time series and the results we obtain apply to this case as well. We take the point of view of generalized linear models theory; see McCullagh and Nelder (1989). The conditional distribution of a categorical time series given its past is multinomial which obviously belongs to the multivariate exponential family of distributions. As such, the theory of generalized linear models can be applied for modeling different types of categorical data; nominal, interval and scale. We will be mostly concerned with nominal data and therefore the multinomial logistic model is the natural candidate for model fitting; see Fahrmeir and Tutz (2001) and Kedem and Fokianos (2002), among other references, for further discussion on modeling issues regarding categorical data. We emphasize that finite state Markov chains provide a simple but prominent model of a categorical time series where lagged values of the response affect the determination of its future states. However, Markov modeling in the context of categorical time series, poses challenging problems. Indeed, as the order of the Markov chain increases so does the number of free parameters; in fact, the number of free parameters increases exponentially fast. Furthermore, the Markovian property requires simultaneous specification of the dynamics of the response and any possible covariates observed jointly; such a specification might not be possible, in general.

We will be studying models for binary and, more generally, categorical time series, which are of infinite order or they are driven by a latent process or a feedback mechanism. This type of models is quite analogous to GARCH models -see Bollerslev (1986)- but they are defined in terms of conditional log–odds instead of conditional variances. In particular, feedback models make possible low dimensional parametrization, yet they can accommodate quite complicated data structures. Examples of feedback models, in the context of binary and categorical time series have been studied recently by Moysiadis and Fokianos (2014) and Fokianos and Moysiadis (2017), among others. We will discuss these results and we will compare them with our findings which improve these works. Models and inference about binary time series, in general, are topics that have been studied by several authors; see Kedem (1980) for an early treatment. Regression modeling, in this context, has been studied by Cox (1981), Stern and Coe (1984), and Slud and Kedem (1994), among others; see also Kedem and Fokianos (2002, Ch. 2-3) for other early references. Recently, binary time series data have been increasingly popular in various financial applications (Breen et al. (1989), Butler and Malaikah (1992), Christoffersen and Diebold (2006), Christoffersen et al. (2007), Startz (2008), Nyberg (2010, 2011, 2013), Kauppi (2012) and Wu and Cui (2014)), but also to other scientific fields. Previous results related to theoretical properties of such models were given by de Jong and Woutersen (2011).

Related work on categorical time series has been reported by Fahrmeir and Kaufmann (1987), Kaufmann (1987), Fokianos and Kedem (2003) and Russell and Engle (1998, 2005) who proposed a categorical time series model for financial transactions data. Alternative classes of models are based on the probit link function. Such autoregressive models have been considered by Zeger and Qaqish (1988), Rydberg and Shephard (2003), Kauppi and Saikkonen (2008), among others. Several other classes of models for the analysis of categorical data have been studied; see the books by Joe (1997) and MacDonald and Zucchini (1997) and the articles by Biswas and Song (2009) and Weiß (2011).

To prove the theoretical results, we will be assuming a contraction type condition; such conditions are usually employed for the theoretical analysis of time series models. For instance, in the case of count time series models, see Doukhan et al. (2012), Fokianos et al. (2009) and Neumann (2011). However, our work is closely related to the modeling approach suggested by Fokianos and Tjøstheim (2011), because the main idea is essentially to employ the so called canonical link process to model the observed data. Note that de Jong and Woutersen (2011) have shown near epoch dependence for a binary time series models but these authors have a different modeling point of view.

Likelihood based inference for the models we study can be developed along the lines of previous references. The proof of consistency and asymptotic normality is based on standard arguments concerning convergence of the score function and the Hessian matrix. However, we mention that this work relaxes considerably previous results. for the case of a model with covariates we improve upon Kaufmann (1987) and Kedem and Fokianos (2002, Ch.3) because we avoid any assumptions on the design of covariates or their boundedness. Since proofs of these facts have been documented in several of the previous references, we do not give any details.

The article is structured as follows: Section 2 discusses general categorical time series models by allowing the conditional probabilities to depend on the whole past of the series. In addition we will be giving a result about the stationarity and ergodicity of chains with complete connections. These results this will be applied to the case of an infinite order autoregressive multinomial logistic model. Section 3 discuss models which might include a latent process. The results obtained by Theorem 1 are applicable to the case of models considered by Moysiadis and Fokianos (2014) and Fokianos and Moysiadis (2017) and improve the previously given stationarity and ergodicity conditions. Finally, Section 4 discuss inclusion of exogenous covariates to the autoregressive multinomial logistic model; Theorem 2 is the main result of this section which discuss existence of such processes and their ergodic properties.

2 Time series autoregressive models for categorical data

2.1 A general approach

Let A be a finite set. For simplicity, we assume that $A = \{1, 2, \dots, N\}$, where N is a nonnegative integer. Suppose that we observe a process with state space A and we are interested on modeling its dynamics. For instance, consider modeling of a stock price change (0 for no change, 1 for positive change and -1 for a negative change; see Russell and Engle (1998)) or sleep state status (see Fokianos and Kedem (2003)). Towards this goal, define a $(N-1)$-dimensional vector $Y_t = (Y_{1t}, Y_{2t}, \dots, Y_{(N-1)t})'$, for $1 \leq t \leq n$, such that

$$Y_{jt} = \begin{cases} 1, & \text{if the } j\text{'th category is observed at time } t, \\ 0, & \text{otherwise,} \end{cases}$$

for all $j = 1, 2, \dots, N-1$. Throughout this work, consider a stochastic processes $(Y_t)_{t\in\mathbb{Z}}$ adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ which is defined through a vector of conditional "success" probabilities, say $p_t \equiv (p_{1t}, p_{2t}, \dots, p_{(N-1)t})'$. In other words

$$p_{jt} = \mathbb{P}\left(Y_{jt} = 1 | \mathcal{F}_{t-1}\right), \quad 1 \leq j \leq N-1. \tag{1}$$

For the last category N, set $Y_{Nt} = 1 - \sum_{j=1}^{N-1} Y_{jt}$ and corresponding success probability $p_{Nt} = 1 - \sum_{j=1}^{N-1} p_{jt}$.

There are several possibilities for autoregressive modeling of processes that take values on a finite space. For instance, assuming that d is a vector and A, B matrices of appropriate dimension, consider the following linear model

$$p_t = d + Ap_{t-1} + BY_{t-1}, \quad t \in \mathbb{Z}, \tag{2}$$

which was studied by Russell and Engle (1998) and Qaqish (2003). Model (2) implies quite complex restrictions on the parameters d, A and B because each element of the vector p_t has to belong in the interval $(0, 1)$. Such restrictions become even more involved when a covariate process is included in (2). To avoid such subtle technicalities, we adapt the generalized linear models point of view; see Fokianos and Kedem (2003) for instance. For $j = 1, 2, \dots, N-1$, define

$$\lambda_{jt} = \log\left(p_{jt}/p_{Nt}\right)$$

and suppose that the vector process $\lambda_t = (\lambda_{1t}, \ldots, \lambda_{(N-1)t})'$ is determined by the infinite order model

$$\lambda_t = g\left(Y_{t-1}, Y_{t-2}, \ldots\right), \tag{3}$$

where g is a suitably defined function. Then, the process $(Y_t)_{t\in\mathbb{Z}}$ which satisfies (1) and (3), takes its values in the set $E = \{e_1, e_2, \ldots, e_{N-1}, \mathbf{0}\}$ where $\{e_1, \ldots, e_{N-1}\}$ is the canonical basis of $\mathbb{R}^{N-1}$ and $\mathbf{0}$ is the null vector of $\mathbb{R}^{N-1}$. Furthermore, $g : E^{\mathbb{N}} \to \mathbb{R}^{N-1}$ is a measurable function and the conditional distribution of Y_t given its past values $Y_{t-1}^{-} \equiv (Y_{t-1}, Y_{t-2}, \ldots)$ possibly depends on its infinite past. A useful example of such process is given by the linear process

$$\lambda_t = d + \sum_{j\geq 1} A_j Y_{t-j} \tag{4}$$

where d is a $(N-1)$-dimensional vector and $(A_j)_{j\geq 1}$ is a sequence of $(N-1)\times(N-1)$ matrices. Comparison of (4) to (2) shows that unnecessary restrictions on the unknown coefficients can be circumvented since the vector $\lambda_t \in \mathbb{R}^{N-1}$. Furthermore, covariates can be easily included in (4) by including an additional additive term. Other categorical type autoregressive models can be considered but (4) has been widely used in several applications. In the case that $N = 2$, then (4) is a simple logistic regression model which has been studied widely in the literature (see Cox and Snell (1970) for an early reference).

Processes, as those we consider in this work, are particular examples of a more general class of processes which are called *chains with complete connections*. Such processes have been widely studied in applied probability; Doeblin and Fortet (1937), Harris (1955) and Iosifescu and Grigorescu (1990). Following the work of Bressaud et al. (1999), we discuss next a coupling technique related to chains with complete connections.

2.2 Some results about chains with complete connection

Throughout this section, consider a finite state space E. For $x, y \in E^{\mathbb{N}}$ and a positive integer m, we write $x \overset{m}{=} y$ if $x_i = y_i$ for $0 \leq i \leq m-1$. Consider a probability kernel $p(\cdot|\cdot)$ defined on $\left(E^{\mathbb{N}}, \mathcal{B}\left(E^{\mathbb{N}}\right)\right)$ and takes values on $(E, \mathcal{B}(E))$ which satisfies the following assumption:

Assumption (A) There exists a sequence $(\gamma_m)_{m\in\mathbb{N}}$ which decreases to zero, as $m \to \infty$, with $\gamma_0 < 1$ and such that for $a \in E$

$$\inf_{x,y:x\overset{m}{=}y} \frac{p\left(a|x\right)}{p\left(a|y\right)} \geq 1 - \gamma_m.$$

A chain with complete connections is a stationary process satisfying Assumption **(A)**.

For $x \in E^{\mathbb{N}}$, consider the chain $(Z_n^x)_{n\in\mathbb{Z}}$ which satisfies that $Z_{-j}^x = x_j$ for $j \geq 1$ and

$$\mathbb{P}\left(Z_n^x = a | Z_{n-j}^x = z_j, j \geq 1\right) = p(a|z) \prod_{j=n+1}^{\infty} \mathbb{1}_{z_j = x_{j-n}}.$$

In addition, given a real-valued sequence $(\gamma_n)_{n\in\mathbb{N}}$, let the Markov chain $\left(S_n^{(\gamma)}\right)_{n\in\mathbb{N}}$ taking values in $\mathbb{N}$ and defined by

$$\mathbb{P}\left(S_0^{(\gamma)} = 0\right) = 1, \quad \mathbb{P}\left(S_{n+1}^{(\gamma)} = i+1 | S_n^{(\gamma)} = i\right) = 1 - \gamma_i, \quad \mathbb{P}\left(S_{n+1}^{(\gamma)} = 0 | S_n^{(\gamma)} = i\right) = \gamma_i.$$

For $n \geq 1$, define the quantity

$$\gamma_n^* = \mathbb{P}\left(S_n^{(\gamma)} = 0\right),$$

which plays a crucial rule for evaluating the mixing coefficients of the chain. The following result is given by Bressaud et al. (1999, Prop. 1 and Lemma 1).

Proposition 1. *For all $x, y \in E^{\mathbb{N}}$, there is a coupling $\left(\left(U_n^{x,y}, V_n^{x,y}\right)\right)_{n\in\mathbb{Z}}$ of $(Z_n^x)_{n\in\mathbb{Z}}$ and $\left(Z_n^y\right)_{n\in\mathbb{Z}}$ such that the integer-valued process $\left(T_n^{x,y}\right)_{n\in\mathbb{Z}}$ defined by*

$$T_n^{x,y} = \inf\left\{m \geq 0 : U_{n-m}^{x,y} \neq V_{n-m}^{x,y}\right\},$$

satisfies

$$\mathbb{P}\left(S_n^{(\gamma)} \geq k\right) \leq \mathbb{P}\left(T_n^{x,y} \geq k\right) \quad \forall k \in \mathbb{N}.$$

Proposition 1 is proved by defining iteratively the pair $\left(U_n^{x,y}, V_n^{x,y}\right)$ using the maximal coupling of the conditional distributions $p\left(\cdot|(u_{n-j})_{j\geq 1}\right)$ and $p\left(\cdot|(v_{n-j})_{j\geq 1}\right)$ (i.e the coupling associated to the total variation distance between these conditional distributions). Proposition 1 yields the following corollary (see also Bressaud et al. (1999, Cor. 1) for a specific case of the following result).

Corollary 1. *For all $k \geq 1$, $x, y \in E^{\mathbb{N}}$ and $B \in \mathcal{B}(E^k)$, we have*

$$\left|\mathbb{P}\left((Z_n^x, \ldots, Z_{n+k}^x) \in B\right) - \mathbb{P}\left((Z_n^y, \ldots, Z_{n+k}^y) \in B\right)\right| \leq \sum_{j=0}^{k} \left(\prod_{m=0}^{j-1} (1 - \gamma_m)\right) \gamma_{n+k-j}^*.$$

Proof. Using Proposition 1, we obtain

$$\begin{aligned}\left|\mathbb{P}\left((Z_n^x, \ldots, Z_{n+k}^x) \in B\right) - \mathbb{P}\left((Z_n^y, \ldots, Z_{n+k}^y) \in B\right)\right| &\leq \mathbb{P}\left((U_n^{x,y}, \ldots, U_{n+k}^{x,y}) \neq (V_n^{x,y}, \ldots, V_{n+k}^{x,y})\right) \\ &\leq \mathbb{P}\left(T_n^{x,y} \leq k\right).\end{aligned}$$

Proposition 1 implies that $\mathbb{P}\left(T_n^{x,y} \leq k\right) \leq \mathbb{P}\left(S_n^{(\gamma)} \leq k\right)$. The result of the corollary now follows by bounding $\mathbb{P}\left(S_n^{(\gamma)} \leq k\right)$ along the lines of the derivation of Bressaud et al. (1999, eq. (4.25)). □

As pointed out in Bressaud et al. (1999), if $\lim_{n\to\infty} \gamma_n^* = 0$, Corollary 1 implies existence and uniqueness of a stationary chain $(Z_n)_{n\in\mathbb{Z}}$ with complete connections and satisfying Assumption **(A)**.

Furthermore, Corollary 1 yields a bound for controlling the $\phi-$mixing coefficients associated with $(Z_n)_{n\in\mathbb{Z}}$. Indeed, recall that for two $\sigma-$algebras $\mathcal{A}$ and $\mathcal{B}$, their $\phi-$mixing coefficients are defined by (see Doukhan (1994), for instance)

$$\phi(\mathcal{A},\mathcal{B}) = \sup_{(A,B)\in\mathcal{A}\times\mathcal{B}:\mathbb{P}(A)>0} \Big(|\mathbb{P}(B|A) - \mathbb{P}(B)|\Big).$$

Then, the ϕ-mixing coefficients of the random sequence $(Z_n)_{n\in\mathbb{Z}}$ are given by

$$\phi(n) = \phi\left(\mathcal{F}_{-\infty,0}, \mathcal{F}_{n,\infty}\right).$$

Since a Borel set on the infinite product can be approximated by a finite union of cylinder sets, we have also

$$\phi(n) = \sup_{k\in\mathbb{N}} \phi\left(\mathcal{F}_{-\infty,0}, \mathcal{F}_{n,n+k}\right).$$

Proposition 2. *Suppose that $\sum_{n\geq 1}\gamma_n^* < \infty$. Then the infinite order stationary Markov chain $(Z_n)_{n\in\mathbb{Z}}$, which exists by Corollary 1, is $\phi-$mixing with mixing coefficients satisfying $\phi(n) \leq \sum_{j\geq n}\gamma_j^*$.*

Proof. Suppose that μ denotes the probability distribution of $(Z_i)_{i\leq -1}$. Then

$$\begin{aligned}\left|\mathbb{P}\left((Z_n^x,\ldots,Z_{n+k}^x)\in B\right) - \mathbb{P}\left((Z_n,\ldots,Z_{n+k})\in B\right)\right| &\leq \int \left|\mathbb{P}\left((Z_n^x,\ldots,Z_{n+k}^x)\in B\right) - \mathbb{P}\left((Z_n^y,\ldots,Z_{n+k}^y)\in B\right)\right|\mu(dy)\\ &\leq \sum_{j=0}^{k}\left(\prod_{m=0}^{j-1}(1-\gamma_m)\right)\gamma_{n+k-j}^*.\end{aligned}$$

But the last bound does not depend on x. Hence, we obtain

$$\begin{aligned}\left|\mathbb{P}\left((Z_n,\ldots,Z_{n+k})\in B|\mathcal{F}_{-\infty,0}\right) - \mathbb{P}\left((Z_n,\ldots,Z_{n+k})\in B\right)\right| &\leq \sum_{j=0}^{k}\left(\prod_{m=0}^{j-1}(1-\gamma_m)\right)\gamma_{n+k-j}^*\\ &\leq \sum_{j\geq n}\gamma_j^*.\end{aligned}$$

The last bound, which does not depend on k and B, is also an upper bound for $\phi(n)$. □

Remark 1. It has been shown in Bressaud et al. (1999, Prop. 2) that

$$\sum_{k\geq 1}\gamma_k < \infty \Rightarrow \sum_{k\geq 1}\gamma_k^* < \infty.$$

Moreover, if $(\gamma_m)_m$ decreases exponentially, then so does $(\gamma^*_n)_n$. Hence, the result of Prop. 2 follows again and if $(\gamma_m)_m$ decreases to zero exponentially fast then so does $(\phi(n))_n$. Note also that the $\phi-$mixing property implies ergodicity of the process; see Bradley (2007, pp. 50-51).

2.3 Application to categorical time series

Recall the categorical time series model $(Y_t)_{t\in\mathbb{Z}}$ whose state space is $E = \{e_1, \ldots, e_{N-1}, \mathbf{0}\}$ and defined by (1) and (3). From the results of the previous subsection, we deduce the following corollary. ($\|\cdot\|$ denotes the Euclidian norm on $\mathbb{R}^{N-1}$.)

Corollary 2. *Assume model (3) and let a function $g : E^{\mathbb{N}} \to \mathbb{R}^{N-1}$ be such that there exist a sequence $(\delta_j)_{j\in\mathbb{N}}$ which satisfies $\sum_{j\in\mathbb{N}} \sum_{k\geq j} \delta_k < \infty$ and*

$$\|g(x) - g(y)\| \leq \sum_{j\in\mathbb{N}} \delta_j \mathbb{1}_{x_j \neq y_j}. \tag{5}$$

Then, there exists a unique stochastic process $(Y_t)_{t\in\mathbb{Z}}$ taking values in E such that

$$\mathbb{P}\left(Y_t = e_j | \mathcal{F}_{t-1}\right) = \frac{\exp\left(g_j\left(Y_{t-1}, Y_{t-2}, \ldots\right)\right)}{1 + \sum_{s=1}^{N-1} \exp\left(g_s\left(Y_{t-1}, Y_{t-2}, \ldots\right)\right)}, \quad 1 \leq j \leq N-1. \tag{6}$$

Moreover $(Y_t)_{t\in\mathbb{Z}}$ is stationary and $\phi-$mixing.

Proof. Denote by $p(\cdot|\cdot)$ the probability kernel defined by

$$p(e_j|x) = F_j\left[g\left(x_0, x_1, \ldots\right)\right], \quad 1 \leq j \leq N-1,$$

where $F_j : \mathbb{R}^{N-1} \to [0, 1]$ is defined for $z \in \mathbb{R}^{N-1}$ by

$$F_j(z) = \frac{\exp(z_j)}{1 + \sum_{s=1}^{N-1} \exp(z_s)}, \quad 1 \leq j \leq N-1.$$

Because of (1), $F_N(z) = \left(1 + \sum_{s=1}^{N-1} \exp(z_s)\right)^{-1}$. The Lipschitz assumption (5) implies that the j'th component of g, say g_j, is bounded, for $j = 1, 2 \ldots, N$. Hence, there exists $\eta > 0$ such that, for all $1 \leq j \leq N-1$ and $x \in E^{\mathbb{N}}$,

$$\eta \leq p(e_j|x), \quad \eta \leq p(\mathbf{0}|x).$$

Moreover, F'_j is bounded, for all j. Set $M = \max_{1\leq j\leq N-1} \sup_{z\in\mathbb{R}^{N-1}} \|F'_j(z)\|$. Then, if $x \overset{m}{=} y$ and $a \in E$, we have that

$$\frac{p(a|x)}{p(a|y)} \geq 1 - \frac{M \sum_{j\geq m+1} \delta_j}{\eta}.$$

Provided that m is large enough, choose $\gamma_m = M\sum_{j\geq m+1}\delta_j/\eta$. Hence, there exists an m such that $\gamma_m = 1-\eta$. Then we have $\sum_{k\geq 1}\gamma_k < \infty$ and using Remark 1, we have also $\sum_{k\geq 1}\gamma_k^* < \infty$. Then from Corollary 1 and Proposition 2, there exists a unique stationary solution $(Y_t)_{t\in\mathbb{Z}}$ satisfying (6) and the solution is $\phi-$mixing. □

We note that the condition $\sum_{j\in\mathbb{N}}\sum_{k\geq j} a_k < \infty$ is equivalent to the condition $\sum_{j\in\mathbb{N}} ja_j < \infty$. For the *infinite order* linear model (4), Corollary 2 applies provided that $\sum_{j\geq 1} j\|A_j\| < \infty$ where $\|A_j\|$ denotes the corresponding operator norm of the matrix A_j. In particular, when $N = 2$, we obtain that the logistic autoregressive model of infinite order is stationary and ϕ-mixing if $\sum_{j\geq 1} j|A_j| < \infty$, where $(A_j)_{j\geq 1}$ belongs to $\mathbb{R}$.

3 Categorical time series with a latent process

In this section, we consider some specific instances of chains with complete connections. Following the methodology of GARCH models (see Engle (1982), Bollerslev (1986) and the book by Francq and Zakoïan (2010) for instance), and recalling the notation introduced in (3) we model the latent process $(\lambda_t)_{t\in\mathbb{Z}}$ to depend additionally on its past values. From a statistical perspective, such parametrization yields parsimony and allows for more flexible structures that can accommodate various forms of autocorrelation. To be specific, suppose that p and q are two positive integers and Let $f : \mathbb{R}^{(N-1)p} \times E^q \to \mathbb{R}^{N-1}$ be a function such that

$$\lambda_t = f\left(\lambda_{t-1},\dots,\lambda_{t-p},Y_{t-1},\dots,Y_{t-q}\right), \quad t\in\mathbb{Z}. \tag{7}$$

We will say that the process $((Y_t,\lambda_t))_{t\in\mathbb{Z}}$ is a solution of the problem $\mathcal{P}_f$ if (7) is satisfied and for each $t\in\mathbb{Z}$, λ_t is $\mathcal{F}_{t-1}-$measurable.

3.1 A general result

For $\underline{y}\in E^q$, define the mapping $G_{\underline{y}} : \mathbb{R}^{(N-1)p} \to \mathbb{R}^{(N-1)p}$ by

$$G_{\underline{y}}(\underline{x}) = \left(f(x_1,\dots,x_p,y_1,\dots,y_q)', x_1',\dots,x_{p-1}'\right)',$$

where $f(\cdot)$ has been defined by (7). The main result of this section is the following.

Theorem 1. *Suppose that there exist an integer $k \geq 1$, $\kappa \in (0,1)$ and $K > 0$ such that*

$$\|G_{\underline{y}}(\underline{x}) - G_{\underline{y}'}(\underline{x}')\| \leq K\left(\|\underline{x} - \underline{x}'\| + \mathbb{1}_{\underline{y} \neq \underline{y}'}\right),$$

and for all $\underline{x}, \underline{x}', \underline{y}_1, \ldots, \underline{y}_q$,

$$\|G_{\underline{y}_1} \circ \cdots \circ G_{\underline{y}_k}(\underline{x}) - G_{\underline{y}_1} \circ \cdots \circ G_{\underline{y}_k}(\underline{x}')\| \leq \kappa \|\underline{x} - \underline{x}'\|.$$

Then, the following hold true:

1. *Let $\underline{x}$ be a vector of $\mathbb{R}^{(N-1)p}$ and $\left(\underline{y}_j\right)_{j\geq 1}$ a sequence of elements of E^q. Then the limit*

$$\lim_{s\to\infty} G_{\underline{y}_1} \circ \cdots \circ G_{\underline{y}_s}(x)$$

   exists and does not depend on x. Let $H : (E^q)^{\mathbb{N}} \to \mathbb{R}^{(N-1)p}$ be the function defined by

$$H\left(\underline{y}_1, \underline{y}_2, \ldots\right) = \lim_{s\to\infty} G_{\underline{y}_1} \circ \cdots \circ G_{\underline{y}_s}(x).$$

   Then the function H is bounded. Moreover there exist $C > 0$ such that

$$\|H\left(\underline{y}_1, \underline{y}_2, \ldots\right) - H\left(\underline{y}'_1, \underline{y}'_2, \ldots\right)\| \leq C \sum_{j\geq 1} \kappa^{j/k} \mathbb{1}_{\underline{y}_j \neq \underline{y}'_j}.$$

2. *A process $((Y_t, \lambda_t))_{t\in\mathbb{Z}}$ is solution of the problem $\mathcal{P}_f$ is and only if $(Y_t)_{t\in\mathbb{Z}}$ is a chain with complete connection associated to a function g (see Corollary 2) defined by*

$$g(Y_{t-1}, Y_{t-2}, \ldots) = H_1(V_t, V_{t-1}, \ldots), \quad V_t = (Y_{t-1}, \ldots, Y_{t-q}).$$

   Here H_1 denotes the $N-1$ first coordiantes of the function H defined previously.

3. *There exists a unique strictly stationary solution to the equations (1) and (7). Moreover the process $(Y_t)_{t\in\mathbb{Z}}$ is $\phi-$mixing with a geometric decrease for the mixing coefficients. This implies the ergodicity of the joint process $((Y_t, \lambda_t))_{t\in\mathbb{Z}}$.*

Proof. 1. The first part of the assertion is a straightforward consequence of the assumption and is omitted. We focus on the proof of the Lipschitz property of the function H. For $j \geq 1$, we set

$$G^{(j)}_{\underline{y}} = G_{\underline{y}_{(j-1)k+1}} \circ G_{\underline{y}_{(j-1)k+2}} \circ \cdots \circ G_{\underline{y}_{jk}}.$$

We have

$$H\left(\underline{y}_1, \underline{y}_2, \ldots\right) = \lim_{s\to\infty} G^{(1)}_{\underline{y}} \circ \cdots \circ G^{(s)}_{\underline{y}}(x).$$

By the stated assumption, we obtain that

$$\|G_{\underline{y}}^{(j)}(x) - G_{\underline{y}'}^{(j)}(x)\| \leq \sum_{\ell=1}^{k} K^\ell \mathbb{1}_{\underline{y}_{(j-1)k+\ell} \neq \underline{y}'_{(j-1)k+\ell}}.$$

Hence

$$\|G_{\underline{y}}^{(1)} \circ \cdots \circ G_{\underline{y}}^{(s)}(x) - G_{\underline{y}'}^{(1)} \circ \cdots \circ G_{\underline{y}'}^{(s)}(x)\| \leq \sum_{j=1}^{\infty} \kappa^j \sum_{\ell=1}^{k} K^\ell \mathbb{1}_{\underline{y}_{(j-1)k+\ell} \neq \underline{y}'_{(j-1)k+\ell}}.$$

By setting $C = (K \vee 1)^k$ and letting $x \to \infty$ we obtain the result.

2. From the first point of the theorem, the necessary condition follows easily. Now let us assume that $\lambda_t = H_1(V_t, V_{t-1}, \ldots)$. Setting $\underline{\lambda}_t = H(V_t, V_{t-1}, \ldots)$, note that the continuity of the function G_{V_t} implies that

$$\begin{aligned} \underline{\lambda}_t &= \lim_{s\to\infty} G_{V_t} \circ G_{V_{t-1}} \circ \cdots G_{V_{t-s}}(x) \\ &= G_{V_t}\left(\lim_{s\to\infty} G_{V_{t-1}} \circ \cdots \circ G_{V_{t-s}}(x)\right) \\ &= G_{V_t}\left(\underline{\lambda}_t\right), \end{aligned}$$

which proofs that $(\lambda_t)_{t\in\mathbb{Z}}$ satisfies (7).

3. The third point is a straightforward consequence of the two first results and of Corollary 2. Moreover the geometric decay of the $\phi-$mixing coefficients has been discussed in the remarks made following Proposition 2. Finally, it is well-known that $\phi-$mixing implies ergodicity of the process $(Y_t)_{t\in\mathbb{Z}}$ and then ergodicity of the process $((\lambda_t, Y_t))_{t\in\mathbb{Z}}$; see Samorodnitsky (2016, Ch. 2), for instance.

□

3.2 Linear models

Let $A_0, A_1, \ldots, A_p, B_1, \ldots, B_q$ be some real matrices of size $(N-1)\times(N-1)$. We assume that

$$f\left(x_1, \ldots, x_p; y_1, \ldots, y_q\right) = A_0 + \sum_{i=1}^{p} A_i x_i + \sum_{i=1}^{q} B_i y_i.$$

Then the above model can be written alternatively as

$$G_{V_t}\left(\underline{x}\right) = \widetilde{A}\underline{x} + B,$$

with

$$\widetilde{A} = \begin{pmatrix} A_1 & \cdots & A_{p-1} & A_p \\ & I_{(N-1)(p-1)} & & 0 \end{pmatrix}, \quad B = \begin{pmatrix} A_0 + \sum_{i=1}^{q} B_i Y_{t-i} \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

where $I_{(N-1)(p-1)}$ denotes the identity matrix of order $(N-1)(p-1)$. Then, the assumptions of Theorem 1 are satisfied if the spectral radius of $\widetilde{A}$ is less than unity (and then the norm of $\widetilde{A}^k$ is less than one if k is large enough) which also means that the roots of the polynomial $\mathcal{P}(z) = \det\left(I_{N-1} - \sum_{i=1}^{p} A_i z^i\right)$ are outside the unit disc; Lütkepohl (2005, Ch.2). For the case $p = q = 1$, this result improves the conditions proved by Moysiadis and Fokianos (2014) since it does not require and additional assumption on the coefficient B_1. In addition, the results answers in affirmative the question posed by Tjøstheim (2012) for the case of binary autoregressive model. Compared with the work of Fokianos and Moysiadis (2017) we note again that for the case of logistic autoregressive modeling with binary data, the obtained conditions simplify considerably since it is only required that the coefficients corresponding to the latent process have sum less than one in absolute value.

3.3 Non-linear models

Recall (7) and assume that there exists a norm $\|\cdot\|$ on $\mathbb{R}^{N-1}$ and some positive real numbers $\beta_1, \ldots, \beta_p, \alpha_1, \ldots, \alpha_q$ with $\alpha = \sum_{i=1}^{p} \alpha_i < 1$ and for all $x_1, x'_1, \ldots, x_p, x'_p, y_1, y'_1, \ldots, y_q, y'_q \in \mathbb{R}^{N-1}$,

$$\|f\left(x_1, \ldots, x_p; y_1, \ldots, y_q\right) - f\left(x'_1, \ldots, x'_p; y'_1, \ldots, y'_q\right)\| \leq \sum_{i=1}^{p} \alpha_i \|x_i - x'_i\| + \sum_{i=1}^{q} \beta_i \|y_i - y'_i\|.$$

It can be proved under this condition that, for a process $(Z_t)_{t\in\mathbb{Z}}$ taking values in E^q, the random mapping

$$G_{Z_t}(x) = \left(f(x, Z_t), x_1, \ldots, x_{p-1}\right)$$

is contracting, after iteration. Indeed, if $x, y \in \mathbb{R}^p$, $\sigma_i^x = x_i$ for $1 \leq i \leq p$ and

$$\sigma_t^x = f\left(\sigma_{t-1}^x, \ldots, \sigma_{t-p}^x, Z_t\right), \quad t \geq p+1,$$

it follows by induction that

$$\left\|\sigma_t^x - \sigma_t^y\right\| \leq \alpha^{\frac{t-p}{p}} \|x - y\|, \quad t \geq p+1.$$

Hence, there exists an integer $m \geq 1$, such that the mapping

$$H_t^{(m)}(x) = G_{Z_t} \circ G_{Z_{t-1}} \circ \cdots \circ G_{Z_{t-m}}(x)$$

satisfies

$$\left\|H_t^{(m)}(x) - H_t^{(m)}(x')\right\| \leq \kappa\|x - x'\|$$

for some $\kappa \in (0, 1)$. Therefore the assumption of Theorem 1 is satisfied. We note again that this condition improves upon the conditions obtained by Moysiadis and Fokianos (2014) and Fokianos and Moysiadis (2017) since they require only that $\alpha < 1$.

4 Inclusion of exogenous covariates

In this section, we study the problem of including a covariate process $(Z_t)_{t\in\mathbb{Z}}$ in an autoregressive categorical time series model. We will be assuming that the covariate process is strongly exogenous. Such an assumption implies that for each time t, Y_t is independent from $(Z_s)_{s\geq t+1}$ conditionally to $Y_{t-1}, Z_t, Y_{t-2}, Z_{t-1}\ldots$ and allows simple computation of the likelihood function. Indeed, if $f\left(\cdot|Y_{t-1}^-, Z_t\right)$ denotes the conditional density of Y_t given $Y_{t-1}, Z_t, Y_{t-2}, Z_{t-1}\ldots$, then, for a bounded measurable function h,

$$\mathbb{E}\left[h\left(Y_1, Y_2, \ldots, Y_n\right)\middle|Z_1 = z_1, \ldots, Z_n = z_n, Y_0^- = y_0^-\right] = \int h(y_1, \ldots, y_n)\prod_{i=1}^n f\left(y_i|y_{i-1}^-, z_i\right)\mu(dy_1)\cdots\mu(dy_n),$$

where μ denotes a reference measure for the model. This type of exogeneity is also called Granger causality or Sims causality in the literature (see, for instance, Gouriéroux and Monfort (1995, Sec. 1.5.2), for a discussion of these different concepts). We will be restricting our study to the case of finite order Markov chains, i.e. the parameter λ_t does not depend on its past values. The general case appears more difficult to tackle and is will be considered in another communication.

Note that even for a finite-state Markov chain, it is difficult to find in the literature a result which guarantees ergodicity when some covariates are included in the dynamic. As we will see, there is an interesting parallel between Markov chains with exogeneous covariates and Markov chains in random environments which were studied in probability theory. In the proof of Theorem 2 given below, we will use an approach discussed in Cogburn (1984) for showing ergodicity of Markov processes in random environments.

4.1 A general result for finite state Markov chains with covariates

We will be discussing results concerning stationarity and ergodicity of a finite state Markov chain which can be jointly observed with a covariate process. In what follows, denote by $Z = (Z_t)_{t\in\mathbb{Z}}$ a stationary process with values in the space $G = \mathbb{R}^d$ and $(Y_t)_{t\in\mathbb{N}}$ a process which takes values in a finite set E. In addition, conditionally on Z, $(Y_t)_{t\in\mathbb{N}}$ is a finite-state inhomogeneous Markov chain. More precisely, we assume that there exist a family of transition matrices $\left\{P_g : g \in G\right\}$ such that

$$\mathbb{P}\left(Y_t = y | Y_{t-1} = x; Z\right) = P_{Z_t}(x, y), \quad (x, y) \in E^2. \tag{8}$$

Throughout the section we will assume the following:

(E1) There exists an integer $m \geq 1$ such that for all $(z_1, z_2, \ldots, z_m, x, y) \in G^m \times E^2$,

$$P_{z_1} P_{z_2} \cdots P_{z_m}(x, y) > 0.$$

(E2) The process Z is mixing in the ergodic theory sense, i.e for all elements A and B of $\mathcal{B}\left(G^{\mathbb{Z}}\right)$, we have

$$\lim_{n\to\infty} \mathbb{P}\left(Z \in A, \tau^n Z \in B\right) = \mathbb{P}(Z \in A)\mathbb{P}(Z \in B),$$

where τ denotes the shift operator on $F^{\mathbb{Z}}$ defined by $\tau Z = \left(Z_{j+1}\right)_{j\in\mathbb{Z}}$.

Note that Assumption (**E1**) implies that a process $(Y_t)_{t\in\mathbb{Z}}$ satisfying (8) also satisfies

$$\mathbb{P}\left(Y_{t+m} = y | Y_t = x, Z\right) > 0, \text{ a.s} \quad (x, y, t) \in E \times E \times \mathbb{Z}.$$

In addition, Assumption (**E2**) is stronger than the assuming ergodicity of the process Z but weaker than the classical strong mixing condition usually employed in the literature. A large number of useful stochastic processes are mixing, for instance the strong mixing processes and Bernoulli shifts defined by $Z_t = H\left(\varepsilon_t, \varepsilon_{t-1}, \ldots\right)$ where H is a measurable function and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an i.i.d sequence. For instance, Samorodnitsky (2016, Ch.2), discusses several properties of the different types of mixing in ergodic theory of stationary processes. Assumption (**E2**) will be employed for obtaining ergodicity for the shift operator τ^m which is not implied by the ergodicity of the shift operator τ. The main result of this section is given by the following theorem.

Theorem 2. *Suppose that (**E1-E2**) hold true. Then there exists a unique stochastic process $(Y_t)_{t\in\mathbb{Z}}$ satisfying (8). Moreover the process $((Y_t, Z_t))_{t\in\mathbb{Z}}$ is ergodic.*

Proof. We first show that the almost sure limit $\lim_{s\to\infty} P_{Z_{t-s}} \cdots P_{Z_t}(x, y)$ exists for each $y \in E$ and does not depend on x. For Markov chains, this condition is comparable to the weak ergodicity notion, but here the limit is taken in the backward sense. See Seneta (2006) for several sufficient conditions ensuring weak ergodicity properties of time-inhomogeneous Markov chains, using ergodicity coefficients. Recall that the so-called Dobrushin's contraction coefficient of a stochastic matrix P is defined by

$$c(P) = \frac{1}{2}\sup_{x\neq y\in E}\|P(x,\cdot)-P(y,\cdot)\|_{TV}.$$

We remind that for two probability mesures μ and ν on the finite set E, the total variation distance between μ and ν is defined by $\|\mu-\nu\|_{TV}=\sum_{x\in E}|\mu(x)-\nu(x)|$. It is well known that we have the contraction

$$\|\mu P-\nu P\|_{TV}\leq c(P)\|\mu-\nu\|_{TV}$$

and for two stochastic matrices P and Q, we have $c(PQ)\leq c(P)c(Q)$.
Moreover $c(P)\leq 1-|E|\min_{x,y\in E}P(x,y)$, where $|E|$ denotes the cardinality of the set E. So Assumption **(E1)** ensures that for all $t\in\mathbb{Z}$, $c\left(P_{Z_t}P_{Z_{t+1}}\cdots P_{Z_{t+m-1}}\right)<1$ a.s. Now let $x\neq x'\in E$, $t\in\mathbb{Z}$ and $s=km+\ell$. we obtain by setting $\rho=1-\eta|E|$,

$$\|P_{Z_{t-s+1}}\cdots P_{Z_t}(x,\cdot)-P_{Z_{t-s+1}}\cdots P_{Z_t}(x',\cdot)\|_{TV}\leq 2c\left(P_{Z_{t-km+1}}\cdots P_{Z_t}\right)\leq 2\prod_{j=0}^{k-1}c\left(P_{Z_{t-(j+1)m+1}}\cdots P_{Z_{t-jm}}\right).$$

From Assumption **(E2)**, the covariate process Z is mixing. Then the process $\left(Z_{t-j}\right)_{j\in\mathbb{Z}}$ is also mixing. Indeed, if θ_t and τ denote the applications defined on $G^{\mathbb{Z}}$ by $\theta_t x=(x_{t-i})_{i\in\mathbb{Z}}$ and $\tau x=(x_{i+1})_{i\in\mathbb{Z}}$ respectively, we have $\tau\circ\theta_t=\theta_t\circ\tau^{-1}$. Then for two borelians $A$and B, we get

$$\begin{aligned}\mathbb{P}\left(\theta_t Z\in A,\tau^n\theta_t Z\in B\right) &= \mathbb{P}\left(Z\in\theta_t^{-1}A,\tau^{-n}Z\in\theta_t^{-1}B\right)\\ &= \mathbb{P}\left(\tau^n Z\in\theta_t^{-1}A,Z\in\theta_t^{-1}B\right)\\ &\to \mathbb{P}\left(\theta_t Z\in A\right)\mathbb{P}\left(\theta_t Z\in B\right).\end{aligned}$$

Moreover, observe that the operator τ^m is ergodic for $\mathbb{P}_Z$. Indeed, if a borelian set A is such that $\tau^m A=A$, we have using assumption **E2**,

$$\mathbb{P}\left(Z\in A\right)=\mathbb{P}\left(\tau^{km}Z\in A,Z\in A\right)\to_{k\to\infty}\mathbb{P}\left(Z\in A\right)^2.$$

Then, we conclude that $\mathbb{P}_Z(A)\in\{0,1\}$, which shows that τ^m is ergodic. Now, using Assumption **(E1)**, we have $\mathbb{E}\log c\left(P_{Z_1}\cdots\mathbb{P}_{Z_m}\right)<0$. Then from the ergodic theorem, we get

$$\prod_{j=0}^{k-1}c\left(P_{Z_{t-(j+1)m+1}}\cdots P_{Z_{t-jm}}\right)=\exp\left(\sum_{j=0}^{k-1}\log c\left(P_{Z_{t-(j+1)m+1}}\cdots P_{Z_{t-jm}}\right)\right)\to_{k\to\infty}0.$$

In addition, when $n \geq s$, we deduce that

$$\|P_{Z_{t-s+1}} \cdots P_{Z_t}(x,\cdot) - P_{Z_{t-n}} \cdots P_{Z_t}(x,\cdot)\|_{TV} \leq 2c\left(P_{Z_{t-s+1}} \cdots P_{Z_t}\right).$$

This shows that the product of matrices $P_{Z_{t-s+1}} \cdots P_{Z_t}$ converges, when $s \to \infty$, to a stochastic matrix whose rows are all equal. Then there exists a measurable function $D : F^{\mathbb{N}} \to E^N$ with $N = |E|$ such that

$$D(Z_t, Z_{t-1}, \ldots) = \lim_{s\to\infty} P_{Z_{t-s+1}} \cdots P_{Z_t}(x,\cdot) \text{ a.s.}$$

Setting $\overline{D}_t = D(Z_t, Z_{t-1}, \ldots)$, $\overline{D}_t$ is a random probability measure on E. For $t \in \mathbb{Z}$, $z \in F^{\mathbb{Z}}$, k a non-negative integer and $y_0, y_1, \ldots, y_k \in E$, we set

$$\mu_{t:t+k}(z; y_0, y_1, \ldots, y_k) = \prod_{i=1}^{k} P_{z_{t+i}}(y_{i-1}, y_i)\overline{D}_t(y_0).$$

From the Kolmogorov extension theorem, there exists for $\mathbb{P}_Z-$almost all values of $z \in \mathcal{B}\left(G^{\mathbb{Z}}\right)$ a unique measure $\mu(z,\cdot)$ on $E^{\mathbb{Z}}$ with marginals $\mu_{t:t+k}(z,\cdot)$. Hence, if ζ denotes the probability distribution of Z, the measure γ defined by

$$\gamma(A \times B) = \int_B \mu(z, A)\zeta(dz), \quad (A, B) \in \mathcal{B}(E^{\mathbb{Z}}) \times \mathcal{B}(F^{\mathbb{Z}}),$$

is that of a couple of stationary process (Y, Z) satisfying (8). To show uniqueness, let $(Y'_t)_{t\in\mathbb{Z}}$ be another stochastic process satisfying (8). Then the distribution of $Y'|Z = z$ is that of a non-homogeneous Markov chain with transitions $\{P_{z_t} : t \in \mathbb{Z}\}$. As shown before, this conditional distribution is unique and equal to $\mu(z,\cdot)$.

Next we show ergodicity of the process $((Y_t, Z_t))_{t\in\mathbb{Z}}$. To this end, we use an approach introduced in Cogburn (1984) for the study of Markov processes in random environment. This type of argument is also used in Sinn and Poupart (2011) for positive transition matrix P_{Z_t} and we give here a more general and shorter proof. The approach used in Cogburn (1984) consists in considering the Markov kernel Q on $E \times F$ defined by

$$Q\left((x, z), \{y\} \times A\right) = P_{z_1}(x, y)\mathbb{1}_A(\tau z), \quad A \in \mathcal{B}\left(F^{\mathbb{Z}}\right).$$

If ν denotes the probability distribution $(Y_t, \tau^t Z)$ which takes values in $E\times F$, then ν is invariant for Q and the process $(H_t)_{t\in\mathbb{Z}}$ defined by $H_t = (Y_t, \tau^t Z)$ is a Markov chain of transition kernel Q. Let C be a $\nu-$ invariant set, i.e $Q((x, z), C) = 1$ for $\nu-$almost every $(x, z) \in C$. Using Corollary 5.11 in Hairer (2006), the Markov chain $(H_t)_{t\in\mathbb{Z}}$ forms an ergodic process if and only if every $\nu-$invariant set C is of measure 0 or 1. In our case, we have $C = \cup_{x\in E}\{x\} \times C_x$ for some $C_x \in \mathcal{B}\left(F^{\mathbb{Z}}\right)$. To this end, we first note that if C is $\nu-$invariant, then

$$\nu(C) = \nu Q(C) = \int_C d\nu(x,z)Q((x,z),C) + \int_{C^c} d\nu(x,z)Q((x,z),C) = \nu(C) + \int_{C^c} d\nu(x,z)Q((x,z),C).$$

Then we get $Q((x,z),C) = 0$ for $\nu-$almost every $(x,z) \in C^c$, the complement of C in $E \times F$. Hence, we obtain $Q((x,z),C) = Q\mathbb{1}_C(x,z) = \mathbb{1}_C(x,z)$ for $\nu-$almost every (x,z). But this also gives $Q^m\mathbb{1}_C = \mathbb{1}_C$, ν a.e, where m has been defined by assumption (**E1**). Moreover, we have that

$$Q^m((x,z),C) = \sum_{y\in E} \mathbb{1}_C(y,\tau^m z)\left[P_{z_1}\cdots P_{z_m}\right](x,y). \tag{9}$$

We write $A = B$ $\nu-$a.e. if $\nu(A\Delta B) = 0$ where $A\Delta B$ denotes the symmetric difference of the sets A and B, i.e $A\Delta B = (A \cap B^c) \cup (A^c \cap B)$. From assumption (**E1**), all the entries of the matrix $P_{z_1}\cdots P_{z_m}$ are positive. Then we deduce that for almost every $(x,z) \in C$, we have $(y,\tau^m z) \in C$ for all $y \in E$. We set $D = \cap_{y\in E} C_y$. Let us denote by ν_1 and ν_2 the marginals of ν. We first remark that for all $A \in \mathcal{B}(F)$, we have

$$\nu(\{x\} \times A) = \int_A \mathbb{P}(Y_0 = x|Z = z)\,\nu_2(dx) = \sum_{y\in E}\int_A \mathbb{P}(Y_0 = x|Y_{-m} = y, Z = z)\,\mathbb{P}(Y_n = y|Z = z)\,\nu_2(dz).$$

Employing again assumption (**E1**), we get $\nu(\{x\} \times A) \geq \eta\nu_2(A)$. For $x \in E$, we set $B_x = \tau^m C_x \setminus D$. As stated above we have

$$\nu(\{(x,z) : z \in B_x\}) = \sum_{x\in E} \nu(\{x\} \times B_x) = 0.$$

We conclude that $\nu_2(B_x) = 0$ for all $x \in E$ and then $\nu_2(\tau^m C_x \setminus C_x) = 0$. By stationarity, $\nu_2(\tau^m C_x) = \nu_2(C_x)$. Therefore, for every $x \in E$, $\tau^m C_x = C_x$, $\mu-$a.e. But using assumption **E2**, we have that

$$\nu_2(C_x) = \nu_2\left(\tau^{km} C_x \cap C_x\right) \to_{k\to\infty} \nu_2(C_x)^2.$$

Then, we conclude that $\nu_2(C_x) \in \{0,1\}$. If $\nu_2(C_x) = 0$ for every x, we easily get $\nu(C) = 0$. Now if there exists $x \in E$ such that $\nu_2(C_x) = 1$, we have, using the equality $\nu_2(B_x) = 0$,

$$1 \leq \nu_2(C_x) = \nu_2(\tau^m C_x) = \nu_2(\tau^m C_x \cap D) \leq \nu_2(D) \leq \min_{y\in E} \nu_2(C_y).$$

Then $\nu_2(C_y) = 1$ for each $y \in E$. Finally we obtain

$$\nu(C) = \sum_{y\in E} \nu\left(\{y\} \times C_y\right) = \sum_{y\in E} \nu_1(y) = 1.$$

Hence, we have shown that the process $(H_t)_{t\in\mathbb{Z}}$ is ergodic and so is the process $((Y_t, Z_t)_{t\in\mathbb{Z}}$.

□

4.2 Application to the multinomial logistic model with covariates

We assume here that conditionally to a covariate process $(Z_t)_{t\in\mathbb{Z}}$ taking values in $\mathbb{R}^d$, the process $(Y_t)_{t\in\mathbb{Z}}$ is a $q-$order Markov chain such that

$$\mathbb{P}\left(Y_t = e_j|Y_{t-1},\ldots,Y_{t-q},Z\right) = \frac{\exp\left(g_j\left(Y_{t-1},\ldots,Y_{t-q};Z_t\right)\right)}{1+\sum_{s=1}^{N-1}\exp\left(g_s\left(Y_{t-1},\ldots,Y_{t-q};Z_t\right)\right)} := Q_{Z_t}\left(Y_{t-q:t-1},e_j\right)$$

for some measurable functions $g_j : E^q\times\mathbb{R}^d$, $1\leq j\leq N-1$. Let us check that assumption **E1** is satisfied for the conditional Markov chain $(X_t)_{t\in\mathbb{Z}}$ defined by $X_t = \left(Y_t', Y_{t-1}',\ldots,Y_{t-q+1}'\right)'$. Conditionally to Z, the process $(X_t)_{t\in\mathbb{Z}}$ defines a time-inhomogeneous Markov chains such that

$$P_{Z_t}\left((u_1,\ldots,u_q),(v_1,\ldots,v_q)\right) : \quad = \quad Q_{Z_t}\left((u_1,\ldots,u_q),v_1\right)\prod_{s=1}^{q-1}\mathbb{1}_{v_{s+1}=u_s}.$$

Since the transition Q_{Z_t} takes only positive values, the assumption **E1** follows by taking $m = q$ and $\gamma = a^q$. Then assuming **E2** for the covariate process, Theorem 2 applies and guarantees the ergodicity of the process $((Y_t,Z_t))_{t\in\mathbb{Z}}$.

References

Biswas, A. and P. X.-K. Song (2009). Discrete-valued ARMA processes. *Statistics & Probability Letters 79*, 1884–1889.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Bradley, R. (2007). *Introduction to Strong Mixing Conditions. Vol 1.* Heber City: UT:Kendrick Press.

Breen, W., L. R. Glosten, and R. Jagannathan (1989). Economic significance of predictable variations in stock index returns. *The Journal of Finance 44*, 1177–1189.

Bressaud, X., R. Fernández, and A. Galves (1999). Decay of correlations for non-Hölderian dynamics. a coupling approach. *Electron. J. Probab. 4*, 1–19.

Butler, K. C. and S. Malaikah (1992). Efficiency and inefficiency in thinly traded stock markets: Kuwait and Saudi Arabia. *Journal of Banking & Finance 16*, 197–210.

Christoffersen, P., F. X. Diebold, R. S. Mariano, A. Tay, and Y. K. Tse (2007). Direction-of-change forecasts for asian equity markets based on conditional variance, skewness and kurtosis dynamics: Evidence from hong kong and singapore. *Journal of Financial Forecasting 1*, 1–22.

Christoffersen, P. F. and F. X. Diebold (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science 52*, 1273–1287.

Cogburn, R. (1984). The ergodic theory of Markov chains in randon environments. *Z. Wahrscheinlichkeits-theory verw. Gebiete 66*, 109–128.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scand. J. Statist. 8*, 93–115.

Cox, D. R. and E. J. Snell (1970). *The Analysis of Binary Data*. London: Chapman & Hall.

de Jong, R. M. and T. Woutersen (2011). Dynamic time series binary choice. *Econometric Theory 27*, 673–702.

Doeblin, W. and R. Fortet (1937). Sur les chaînes à liaisons complètes. *Bull. Soc. Math. France 65*, 132–148.

Doukhan, P. (1994). *Mixing: properties and examples*. Number 85 in Lecture Notes in Statistics. New York: Springer-Verlag.

Doukhan, P., K. Fokianos, and D. Tjøstheim (2012). On weak dependence conditions for Poisson autoregressions. *Statist. Probab. Lett. 82*, 942–948.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Fahrmeir, L. and H. Kaufmann (1987). Regression models for nonstationary categorical time series. *Journal of Time Series Analysis 8*, 147–160.

Fahrmeir, L. and G. Tutz (2001). *Multivariate statistical modelling based on generalized linear models* (Second ed.). Springer Series in Statistics. New York: Springer-Verlag. With contributions by Wolfgang Hennevogl.

Fokianos, K. and B. Kedem (2003). Regression theory for categorical time series. *Statist. Sci. 18*, 357–376.

Fokianos, K. and T. Moysiadis (2017). Binary time series friven by a latent process. *Econometrics and Statistics 2*, 117–130.

Fokianos, K., A. Rahbek, and D. Tjostheim (2009). Poisson autoregression. *J. Amer. Statist. Assoc. 104*, 1430–1439.

Fokianos, K. and D. Tjøstheim (2011). Log-linear Poisson autoregression. *J. Multivariate Anal. 102*, 563–578.

Francq, C. and J.-M. Zakoïan (2010). *GARCH models: Stracture, Statistical Inference and Financial Applications*. United Kingdom: Wiley.

Gouriéroux, C. and A. Monfort (1995). *Statistics and Econometric Models. Volume 1*. Cambridge University Press.

Hairer, M. (2006). *Ergodic properties of Markov processes*. Lecture notes available at http://www.hairer.org/notes/Markov.pdf.

Harris, T. (1955). On chains of infinite order. *Pacific J. Math. 5*, 707–724.

Iosifescu, M. and S. Grigorescu (1990). *Dependence with Complete Connections and its Applications*. Cambridge University Press.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Kaufmann, H. (1987). Regression models for nonstationary categorical time series: Asymptotic estimation theory. *Annals of Statistics 15*, 79–98.

Kauppi, H. (2012). Predicting the direction of the Fed's target rate. *Journal of Forecasting 31*, 47–67.

Kauppi, H. and P. Saikkonen (2008). Predicting US recessions with dynamic binary response models. *The Review of Economics and Statistics 90*, 777–791.

Kedem, B. (1980). *Binary Time Series*. Marcel Dekker, New York.

Kedem, B. and K. Fokianos (2002). *Regression models for time series analysis*. Hoboken, NJ: Wiley.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis* (1st ed.). Berlin: Springer.

MacDonald, I. L. and W. Zucchini (1997). *Hidden Markov and Other Models for Discrete–valued Time Series*. London: Chapman & Hall.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (2nd ed.). London: Chapman & Hall.

Moysiadis, T. and K. Fokianos (2014). On binary and categorical time series models with feedback. *J. Multivariate Anal. 131*, 209–228.

Neumann, M. (2011). Absolute regularity and ergodicity of poisson count processes. *Bernoulli 17*, 1268–1284.

Nyberg, H. (2010). Dynamic probit models and financial variables in recession forecasting. *Journal of Forecasting 29*, 215–230.

Nyberg, H. (2011). Forecasting the direction of the us stock market with dynamic binary probit models. *International Journal of Forecasting 27*, 561–578.

Nyberg, H. (2013). Predicting bear and bull stock markets with dynamic binary time series models. *Journal of Banking & Finance 37*, 3351–3363.

Qaqish, B. F. (2003). A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika 90*, 455–463.

Russell, J. R. and R. F. Engle (1998). Econometric Analysis of Discrete-Valued Irregularly-Spaced Financial Transactions Data Using a New Autoregressive Conditional Multinomial Model. *SSRN eLibrary*.

Russell, J. R. and R. F. Engle (2005). A discrete-state continuous-time model of financial transactions prices and times. *Journal of Business and Economic Statistics 23*, 166–180.

Rydberg, T. H. and N. Shephard (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics 1*, 2–25.

Samorodnitsky, G. (2016). *Stochastic Processes and Long Range Dependence*. Springer.

Seneta, E. (2006). *Non-negative matrices and Markov chains*. Springer Series in Statistics. Springer, New York. Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544].

Sinn, M. and P. Poupart (2011). Asymptotic theory for linear-chain conditional random fields. In G. Gordon, D. Dunson, and M. Dudik (Eds.), *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Volume 15 of *Proceedings of Machine Learning Research*, Fort Lauderdale, FL, USA, pp. 679–687.

Slud, E. V. and B. Kedem (1994). Partial likelihood analysis of logistic regression and autoregression. *Statist. Sinica 4*, 89–106.

Startz, R. (2008). Binomial autoregressive moving average models with an application to US recessions. *Journal of Business & Economic statistics 26*, 1–8.

Stern, R. D. and R. Coe (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society. Series A (General) 147*, pp. 1–34.

Tjøstheim, D. (2012). Rejoinder on: Some recent theory for autoregressive count time series. *TEST 21*, 469–476.

Weiß, C. (2011). Generalized choice models for categorical time series. *J. Statist. Plann. Inference 141*, 2849–2862.

Wu, R. and Y. Cui (2014). A parameter-driven logit regression model for binary time series. *Journal of Time Series Analysis 35*, 462–477.

Zeger, S. L. and B. Qaqish (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics 44*, 1019–1031.