

Coupling and perturbation techniques for categorical time series

LIONEL TRUQUET

UMR 9194 CNRS CREST, ENSAI, Campus de Ker-Lann, Rue Blaise Pascal – BP 37203, 35712 BRUZ cedex, France. E-mail: lionel.truquet@ensai.fr

We present a general approach for studying autoregressive categorical time series models with dependence of infinite order and defined conditional on an exogenous covariate process. To this end, we adapt a coupling approach, developed in the literature for bounding the relaxation speed of a chain with complete connections and from which we derive a perturbation result for non-homogenous versions of such chains. We then study stationarity, ergodicity and dependence properties of some chains with complete connections and exogenous covariates. As a consequence, we obtain a general framework for studying some observation-driven time series models used both in statistics and econometrics but without theoretical support.

Keywords: categorical times series; chains with complete connections; coupling; dependence properties

1. Introduction

Categorical time series are widely encountered in various fields. For instance, in climate analysis, Guanche et al. [24] studied the dynamic of weather types, Hao et al. [25] the prediction of drought periods. In finance, Russell and Engle [41] or Rydberg and Shephard [42] studied the dynamic of price movements. In economics, Kauppi and Saikkonen [29] consider the prediction of recession periods. Several type of models used for modeling categorical time series can be found in the survey of Fokianos and Kedem [20]. Though lots of time series models have been developed in the literature, it is difficult to find a general framework for which inclusion of exogenous covariates is mathematically justified. This is one of the important differences between the theoretical results found in time series analysis and the models used by the practitioners which most of the time, are based on exogenous covariates. A notable exception is the contribution of Kaufmann [28] who considered estimation in autoregressive logistic type models when deterministic regressors are included in the dynamic. More recently, Fokianos and Truquet [22] considered general Markov models with random covariates. However, most of the categorical time series models used in practice are “observation-driven” (see below for a definition), especially in econometrics. Fokianos and Truquet [22] also considered this class of non-Markovian processes but without covariates and it seems that a general approach for studying a wide class of categorical time series models with exogenous covariates is still not available. In this paper, we provide such a framework by using a formalism introduced for studying a general class of finite-state stochastic processes, the chains with complete connections. These processes, initially considered by Doeblin and Fortet [15], have an interest in probability theory, statistical mechanics or ergodic theory. See in particular Harris [26], Iosifescu and Grigorescu [27], Bressaud et al. [1,2], Fernandez and Galves [18] and Comets et al. [8] for many of their theoretical properties.

Chains with complete connections also contain stochastic chains with memory of variable length as a special case, the latter class, initially introduced by Rissanen [40] for data compression, has also applications in linguistic, see Galves et al. [23] or for protein classification, see, for instance, Busch et al. [3].

In this paper, we also consider such chains with complete connections but defined conditional on a covariate process. More precisely, we want to study stochastic processes $(Y_t)_{t \in \mathbb{Z}}$ defined by

$$\mathbb{P}(Y_t = y | Y_{t-1}^-, X_t^-) = q(y | Y_{t-1}^-, X_t^-), \quad y \in E, \quad (1)$$

where $(X_t)_{t \in \mathbb{Z}}$ is a covariate process taking values in \mathbb{R}^d , E is a finite set and q is a transition kernel. We will extensively use the notation $x_t^- = (x_t, x_{t-1}, \dots)$ for a sequence $(x_t)_{t \in \mathbb{Z}}$. Without additional assumptions on the two processes X and Y , (1) is difficult to study theoretically. We will assume further that

$$\mathbb{P}(Y_t = y | Y_{t-1}^-, X) = \mathbb{P}(Y_t = y | Y_{t-1}^-, X_t^-), \quad X := (X_t)_{t \in \mathbb{Z}}. \quad (2)$$

If condition (2) is satisfied, $(Y_t)_{t \in \mathbb{Z}}$ is, conditional on X , a time-inhomogenous chain with complete connections and transition kernels $\{q(\cdot | \cdot, X_t^-) : t \in \mathbb{Z}\}$. Condition (2) also means that Y_t is independent of $(X_{t+1}, X_{t+2}, \dots)$ conditional on $((Y_{j-1}, X_j))_{j \leq t}$. In econometrics, the latter conditional independence assumptions is called strict exogeneity. Initially introduced by Sims [43] for linear models, the concept of strict exogeneity was extended by Chamberlain [5] to categorical time series. Chamberlain [5] also showed that under additional regularity conditions, strict exogeneity is equivalent to non-Granger causality, which means that X_{t+1} is independent of Y_t, Y_{t-1}, \dots , conditional on X_t, X_{t-1}, \dots . This roughly means that the covariate process evolves in a totally autonomous way and that, given all the information available up to time t , past values of the outcome will not influence future values of the covariates. Let us also mention that such strict exogeneity condition is a standard assumption in Markov-switching models, for which the dynamic of the time series under study is defined conditional on an unobserved Markov chain X . In probability theory, this exogeneity notion appears implicitly in the literature of stochastic processes in random environments. Finite-state Markov chains in random environments are a particular case of stochastic processes satisfying (1) and (2). They are studied for instance, in Cogburn [7] and Kifer [30] but no result seems to be available for chains with complete connections. Strict exogeneity has of course some limitations for time series analysis, it is a rather strong assumption. However, it is easier to formulate a general theory in this context, other conditional distributions such as $X_{t+1} | Y_t^-, X_t^-$ need not to be specified.

Stochastic processes defined by (1) are of course of theoretical interest but for applications to time series analysis, one of the challenging problem is to find parsimonious versions of (1). One important class of models are called observation-driven, following the classification proposed by Cox et al. [9]. For model (1), an observation-driven model is obtained assuming that $q(\cdot | Y_{t-1}, X_t^-) = q(\cdot | \mu_t)$ with

$$\mu_t = G(\mu_{t-1}, \dots, \mu_{t-q}, Y_{t-1}, \dots, Y_{t-p}, X_t). \quad (3)$$

Without exogenous covariates, observation-driven models were widely studied, in particular for count time series. See in particular Fokianos et al. [21], Neumann [36], Woodard et al. [46], Douc

et al. [16]. These models are mainly studied using Markov chain techniques due to the Markov properties of the process (Y_t, μ_t) . However, as pointed out in Woodard et al. [46] or Douc et al. [16], for discrete time series, such Markov chains do not satisfy irreducibility properties. In particular the latent variable μ_t is not discrete and not necessarily absolutely continuous. More sophisticated techniques have then been developed to study existence of stationary distributions. Such contributions are often limited to the case $p = q = 1$ and do not consider the problem of exogenous covariates. In contrast, for the special case of categorical time series, one can develop a much more general approach, considering observation-driven models as a particular case of infinite dependence. This approach was recently used by Fokianos and Truquet [22]. However, inclusion of exogenous covariates is a more tricky problem and has not been considered before for model (3) or (1). More generally, despite its fundamental importance for practical applications, the problem of covariates inclusion is often ignored in the time series literature, except for linear models. In Section 4, we make a review of many observation-driven models proposed in econometrics for the study of categorical time series and that can be studied under our general framework.

A crucial point for studying our models is to control how fast the process (Y_t) in (1) loses memory of its initial values. For homogenous chains, Bressaud et al. [1] developed a nice result based on the maximal coupling. We will adapt their result to our context, which will be crucial for defining our models and studying many of their properties.

Another important problem addressed in this paper concerns dependence properties of the process, which are essential to control the behavior of partial sums. While chains with complete connections satisfies ϕ -mixing properties under rather general assumptions (see Fokianos and Truquet [22]), finding dependence properties for the joint process (Y_t, X_t) in (1) is quite challenging. For the example of observation-driven models, such properties are important to control the behavior of partial sums of type $\frac{1}{n} \sum_{t=1}^n f(Y_t^-, X_t^-, \mu_t^-)$. Our aim here is to avoid to prove the so-called “quenched” results found in the literature of processes in random environments, see, for instance, Kifer [31] for a central limit theorem of this type for Markov chains in random environments, and which consist in studying the limiting behavior of the partial sums conditional on the environment $X(\omega)$. In contrast in this paper, we explain how to get β -mixing properties and τ -dependence (see Section 5 for a definition) for this joint process, leading the possibility to use various invariance principle and deviation inequalities for the aforementioned partial sums. To this end, we will use a coupling approach. If a “good” coupling for the covariate process X exists, one can define a coupling of Y conditional on X , with two paths having different transition kernels that will be adjacent at infinity. This is why we will derive in Section 2 a perturbation result for chains with complete connections, obtained via coupling. Such a result also has of independent interest.

The paper is organized as follows. In Section 2, we state a general result for non-homogenous chain with complete connections. In particular, we generalize a result of Bressaud et al. [1] for controlling the relaxation speed of such chains and we also compare the dynamic of two such chains possessing different transition kernels. In Section 3, we give some conditions on the transition kernel $q(\cdot|\cdot)$ that guaranty existence and uniqueness of a stationary and ergodic solution for the problem (1). Many examples are given in Section 4, with a detailed treatment of some observation-driven models used in the econometric literature. Section 5 is devoted to the dependence properties of the solution, absolute regularity or τ -dependence. We mention several

possible applications of our results in statistics in Section 6. The proofs of our results are postponed to Section 7. Finally, an auxiliary lemma for observation-driven models is given in an Appendix. Proofs of some of our results can be found in the supplementary material Truquet [44].

2. Perturbation of chains with complete connections

We denote by \mathbb{N} the set of natural integers $\{0, 1, \dots\}$ and $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$. For a finite set F , we will denote by $\mathcal{P}(F)$ is the set of all subsets of F . Moreover, if ν_1 and ν_2 are two probability measures on F , the total variation distance between ν_1 and ν_2 is defined by

$$d_{TV}(\nu_1, \nu_2) = \frac{1}{2} \sum_{f \in F} |\nu_1(f) - \nu_2(f)|.$$

We remind the following expression

$$d_{TV}(\nu_1, \nu_2) = \inf\{\mathbb{P}(U \neq V) : U \sim \nu_1, V \sim \nu_2\}.$$

Moreover one can define two random variables U and V such that the infimum is attained. A such pair (U, V) is called a maximal coupling, since its distribution maximizes the probability of the diagonal. See den Hollander [14], page 15 for an explicit construction of such coupling and for a proof of the previous equality. For $y, \bar{y} \in E^{\mathbb{N}}$ and a positive integer m , we write $y \stackrel{m}{=} \bar{y}$ if $y_i = \bar{y}_i$ for $0 \leq i \leq m - 1$. Finally we denote by $\mathcal{B}(\mathcal{X})$ the borel sigma-algebra associated to a topological space \mathcal{X} .

2.1. A general result

Throughout the section, we will denote by E a finite set. Let also \mathcal{X} be a Polish space. For any $x \in \mathcal{X}$, we consider two sequences $(q_t^x)_{t \in \mathbb{Z}}$ and $(\bar{q}_t^x)_{t \in \mathbb{Z}}$ of probability kernels from $(E^{\mathbb{N}}, \mathcal{P}(E)^{\otimes \mathbb{N}})$ to $(E, \mathcal{P}(E))$. For our applications to time series, the case $q_t^x = q(\cdot | x_t^-)$ will be of interest. The two following assumptions will be needed.

A1. The applications $(y, z, x) \mapsto q_t^x(y|z)$ and $(y, z, x) \mapsto \bar{q}_t^x(y|z)$ are measurable and take positive values.

A2. Setting

$$b_m := \sup_{t \in \mathbb{Z}} \sup_{x \in \mathcal{X}} \sup_{y \stackrel{m}{=} \bar{y}} d_{TV}(q_t^x(\cdot|y), q_t^x(\cdot|\bar{y})),$$

we have $b_0 < 1$ and $\lim_{m \rightarrow \infty} b_m = 0$.

Let us now introduce some additional notations. In what follows, we fix $t_0 \in \mathbb{Z}$. For $z \in E^{\mathbb{N}}$ and $x \in \mathcal{X}$, we denote by $Q_{t_0,x,z}$ the probability distribution on $(E^{\mathbb{N}}, \mathcal{P}(E)^{\otimes \mathbb{N}})$ defined by

$$Q_{t_0,x,z} \left(\prod_{i=1}^n \{y_i\} \times \prod_{i=n+1}^{\infty} E \right) = \prod_{i=1}^n q_{t_0+i}^x(y_i | y_{i-1}^-)$$

with the convention $y_{-j} = z_j$ for $j \geq 0$. The measure $Q_{t_0,x,z}$ coincides with the joint distribution of $Y_{t_0+1}, Y_{t_0+2}, \dots$ given $Y_{t_0}, Y_{t_0-1}, \dots$ for a chain $(Y_t)_{t \in \mathbb{Z}}$ defined from the sequence of transition kernels $(q_t^x)_{t \in \mathbb{Z}}$. We define $\bar{Q}_{t_0,x,z}$ in the same way, replacing the transition kernels q_t^x with \bar{q}_t^x in the previous expression.

Lemma 1. *Assume that Assumptions A1–A2 hold true. Then for any $x \in \mathcal{X}$ and any couple $(z, \bar{z}) \in E^{\mathbb{N}} \times E^{\mathbb{N}}$, there exists a probability measure $\tilde{Q}_{t_0,x,z,\bar{z}}$ on $(E^{\mathbb{N}^*} \times E^{\mathbb{N}^*}, \mathcal{P}(E)^{\otimes \mathbb{N}^*} \otimes \mathcal{P}(E)^{\otimes \mathbb{N}^*})$ such that the three following conditions are satisfied.*

1. For $A, B \in \mathcal{P}(E)^{\otimes \mathbb{N}^*}$, we have

$$\tilde{Q}_{t_0,x,z,\bar{z}}(A \times E^{\mathbb{N}^*}) = Q_{t_0,x,z}(A), \quad \tilde{Q}_{t_0,x,z,\bar{z}}(E^{\mathbb{N}^*} \times B) = \bar{Q}_{t_0,x,\bar{z}}(B). \tag{4}$$

2. For $t \geq 1$,

$$\begin{aligned} & \tilde{Q}_{t_0,x,z,\bar{z}}(\{(y, \bar{y}) \in E^{\mathbb{N}^*} \times E^{\mathbb{N}^*} : y_t \neq \bar{y}_t\}) \\ & \leq b_{t-1}^* + \sup_{s \in E^{\mathbb{N}}} d_{\text{TV}}(q_{t+t_0}^x(\cdot | s), \bar{q}_{t+t_0}^x(\cdot | s)) \\ & \quad + \sum_{\ell=0}^{t-2} b_{\ell}^* \sup_{s \in E^{\mathbb{N}}} d_{\text{TV}}(q_{t+t_0-\ell-1}^x(\cdot | s), \bar{q}_{t+t_0-\ell-1}^x(\cdot | s)), \end{aligned} \tag{5}$$

where $b_0^* = b_0$ and for $n \geq 1$, b_n^* is equal to $\mathbb{P}(S_n^{(b)} = 0)$ where $(S_n^{(b)})_{n \geq 0}$ is a time-homogeneous Markov chain, starting at 0 and with transition matrix P defined by

$$P(i, i + 1) = 1 - b_i, \quad P(i, 0) = b_i, \quad i \in \mathbb{N}.$$

3. For all $C \in \mathcal{P}(E)^{\otimes \mathbb{N}^*} \otimes \mathcal{P}(E)^{\otimes \mathbb{N}^*}$, the mapping $(x, z, \bar{z}) \mapsto \tilde{Q}_{t_0,x,z,\bar{z}}(C)$ is measurable as a mapping from $(\mathcal{X} \times E^{\mathbb{N}} \times E^{\mathbb{N}}, \mathcal{B}(\mathcal{X}) \otimes \mathcal{P}(E)^{\otimes \mathbb{N}} \otimes \mathcal{P}(E)^{\otimes \mathbb{N}})$ to $([0, 1], \mathcal{B}([0, 1]))$.

Notes.

1. From Lemma 1, we get a coupling of two chains defined from different transition kernels q_t^x or \bar{q}_t^x and initialized with arbitrary past sequences. A control of the total variation distance between the finite dimensional distributions of two such chains can be deduced. For $i \in \mathbb{N}^*$, we denote by y_i (resp. \bar{y}_i) the mapping from $E^{\mathbb{N}^*} \times E^{\mathbb{N}^*}$ to E defined by $y_i(w, \bar{w}) = w_i$ (resp. $\bar{y}_i(w, \bar{w}) = \bar{w}_i$), $w, \bar{w} \in E^{\mathbb{N}^*}$. Let $1 \leq s \leq \ell$. If $Q_{t_0,x,z}^{(s,\ell)}$ and $\bar{Q}_{t_0,x,z}^{(s,\ell)}$ denote the

restriction of $Q_{t_0,x,z}$ (resp. $\bar{Q}_{t_0,x,z}$) to $\sigma(y_i : s \leq i \leq \ell)$, we have

$$d_{TV}(Q_{t_0,x,z}^{(s,\ell)}, \bar{Q}_{t_0,x,z}^{(s,\ell)}) \leq \tilde{Q}_{t_0,x,z,\bar{z}}(y_t \neq \bar{y}_t; \text{ for some } s \leq t \leq \ell) \leq \sum_{t=s}^{\ell} \tilde{Q}_{t_0,x,z,\bar{z}}(y_t \neq \bar{y}_t) \tag{6}$$

and the total variation distance can be then bounded from (5).

- When the $q_t^x \equiv \bar{q}_t^x \equiv q$ and setting $\tilde{Q}_{t_0,x,z,\bar{z}} = \tilde{Q}_{t_0,z,\bar{z}}$, Lemma 1 shows that

$$\tilde{Q}_{t_0,z,\bar{z}}(y_t \neq \bar{y}_t) \leq b_{t-t_0-1}^*$$

and we simply get control of the total variation distance between the marginals at time t , when a time-homogeneous chain with complete connections is initialized with two different sequences. Such result has been proved by Bressaud et al. [1] under a log-continuity assumption for the transition kernel q . Since we use an assumption slightly weaker in S2, we will rewrite a detailed proof for the previous bound using our assumptions.

- Assumption A2 implies a uniform control for the total variation distance between two kernels having their first m -past values equal. This total variation distance decreases to 0 when $m \rightarrow 0$. In the literature of homogenous chains with complete connections, the coefficient b_m is called the variation of order m of the kernel. See Fernandez and Maillard [19] for a discussion about the role of these coefficients for proving uniqueness properties of stationary measures associated to the transition kernel. The decrease of these coefficients measure the loss of memory with respect to values far away in the past.

Next, we provide a perturbation result for homogeneous chains with complete connections, that is, $q_t \equiv q$ and $\bar{q}_t \equiv \bar{q}$ for all integer t . This result will not be used in the rest of the paper. However, it extends a standard perturbation result for finite-state Markov chains and has then an independent interest. If $\sum_{m \geq 1} b_m < \infty$ then $\sum_{m \geq 1} b_m^* < \infty$ (see the supplementary material, Lemma 4). In this case, $b_m^* \rightarrow 0$ and there exists a stationary chain $(Y_k)_{k \in \mathbb{Z}}$ with complete connection and transition kernel q . Moreover, the stationary distribution of such chain is unique. Existence and unicity can hold under slightly weaker conditions that will not be discussed here. See Bressaud et al. [1], Remark 1. We will also assume that the transition kernel \bar{q} satisfies Assumptions A1–A2 with summable coefficients \bar{b}_m . By setting $t_0 = 0$ and letting t going to $-\infty$ in Lemma 1, we obtain the following result.

Corollary 1. *Assume that $\sum_{m \geq 1} (b_m + \bar{b}_m) < \infty$ and let π ($\bar{\pi}$ resp.) be the marginal distribution of the chain with transition kernel q (\bar{q} resp.). Then*

$$d_{TV}(\pi, \bar{\pi}) \leq \left(1 + \sum_{m \geq 0} b_m^*\right) \cdot \sup_{y \in E^{\mathbb{N}}} d_{TV}(q(\cdot|y), \bar{q}(\cdot|y)).$$

Corollary 1 shows that the marginal distribution of the chain is a Lipschitz functional of its transition kernel. Let us specialize this result to the Markov case, that is, $q(\cdot|\cdot)$ is a stochastic matrix on E . In this case, $b_m = 0$ for $m \geq 1$ and it is easily seen that $b_m^* = b_0^m$. We obtain

$1 + \sum_{m \geq 0} b_m^* = (1 - b_0)^{-1}$. We then recover a basic result for the perturbation of Markov chain using the ergodicity coefficient b_0 of the Markov chain with transition q . See, for instance, Mitrophanov [35], Theorem 3.2.

3. Stationary categorical time series models with covariates

In this section, we consider a finite set E with cardinality N . We will consider a stationary covariate process $X = (X_t)_{t \in \mathbb{Z}}$ taking values in $(\mathbb{R}^d, |\cdot|)$ where $|\cdot|$ is a norm on \mathbb{R}^d . Let $(Y_t)_{t \in \mathbb{Z}}$ a time series taking values in E and such that

$$\mathbb{P}(Y_t = w | Y_{t-1}^-, X) = q(w | Y_{t-1}^-, X_t^-), \quad t \in \mathbb{Z}. \tag{7}$$

We assume that for every $t \in \mathbb{Z}$, the mapping $(w, y, x) \mapsto q(w | y, x_t^-)$ is measurable, as a mapping from $E \times E^{\mathbb{N}} \times \mathcal{D}$ to $(0, 1)$, where $\mathcal{D} \in \mathcal{B}(\mathbb{R}^d)^{\otimes \mathbb{Z}}$ is such that $\mathbb{P}(X \in \mathcal{D}) = 1$. Moreover, we impose $\sum_{w \in E} q(w | y, x) = 1$ for all $(y, x) \in E^{\mathbb{N}} \times \mathcal{D}$.

3.1. Existence of a stationary and ergodic solution

The following assumptions will be needed.

S1. The covariate process $X = (X_t)_{t \in \mathbb{Z}}$ stationary and ergodic.

S2. Setting for $m \geq 0$,

$$b_m = \sup \{ d_{TV}(q(\cdot | y, x_t^-), q(\cdot | y', x_t^-)) : (y, y', x) \in E^{\mathbb{N}} \times E^{\mathbb{N}} \times \mathcal{D}, t \in \mathbb{Z}, y \stackrel{m}{=} y' \},$$

we have $b_0 < 1$ and $\sum_{m \geq 0} b_m < \infty$.

Note. Assumption **S2** guarantees that $\sum_{m \geq 0} b_m^* < \infty$, where the b_m^* 's are related to the b'_m 's as described in Lemma 1. A proof can be found in Bressaud et al. [1], Proposition 2. See also the supplementary material, Lemma 4. Basically, the decrease of the sequence $(b_m^*)_{m \geq 0}$ is of the same order as the sequence $(b_m)_{m \geq 0}$. It should be noted that we impose a control of the total variation distances which is uniform with respect to the path of the covariate process X . We did not find a solution for removing this assumption. However, as we will see in the examples, when the contribution of the covariates is additive in some generalized linear models, this assumption is often satisfied even if the covariate process is unbounded.

Theorem 1. Assume that the Assumptions **S1**–**S2** hold true.

1. There then exists a stochastic processes $(Y_t)_{t \in \mathbb{Z}}$ satisfying (7). The probability distribution of the pair (Y, X) is unique. Moreover for any bounded measurable function $h : E^{\mathbb{N}} \rightarrow \mathbb{R}$, we have

$$\mathbb{E}[h(Y_t^-) | X] = \mathbb{E}[h(Y_t^-) | X_t^-].$$

2. The bivariate process $((Y_t, X_t))_{t \in \mathbb{Z}}$ is stationary and ergodic.

Note. As shown in Section 4, Theorem 1 can be applied to many examples of categorical time series models found in the literature. In its present form, such a result cannot be applied to countable infinite state spaces. In the infinite countable case, existence and uniqueness of stationary measures for classical chains with complete connections is more difficult to get. In particular, a standard uniqueness criterion is based on some oscillation coefficients that differ from the b'_m s coefficients and Dobrushin’s contraction condition. See the recent contribution of Chazottes et al. [6] and the references therein for a discussion. For infinite state spaces, Dobrushin’s uniqueness condition generally imposes serious restrictions on the transition kernel q . Moreover, such a condition is not necessarily satisfied by the examples given in the next section.

4. Examples

We now provide many examples of categorical time series models satisfying our assumptions. We study in particular some observation-driven models proposed in the literature, which are parsimonious and of interest for applications in statistics.

4.1. Generalized linear model for binary time series

Here we assume that $E = \{0, 1\}$. We consider the following binary time series model defined by

$$\mathbb{P}(Y_t = 1 | Y_{t-1}^-, X) = F(\mu_t), \quad \mu_t = \sum_{j=1}^{\infty} a_j Y_{t-j} + \gamma' X_t, \tag{8}$$

where F is a cumulative distribution function, $(a_j)_{j \geq 1}$ is a summable sequence of real numbers and $\gamma \in \mathbb{R}^d$. Model (8) extends the model considered by Comets et al. [8] which does not contain exogenous regressors.

Proposition 1. *Assume that F is Lipschitz, takes values in $(0, 1)$, $\sum_{j \geq 1} j|a_j| < \infty$ and X satisfies Assumption S1. There then exists a unique stationary process $(Y_t)_{t \in \mathbb{Z}}$ satisfying (8). Moreover, the bivariate process $((Y_t, X_t))_{t \in \mathbb{Z}}$ is stationary and ergodic.*

Model (9) is of theoretical interest but in practice observation-driven models lead to parsimonious representations of such dynamics. Let us consider the following version.

$$\mathbb{P}(Y_t = 1 | Y_{t-1}^-, X) = F(\mu_t), \quad \mu_t = \sum_{j=1}^q \beta_j \mu_{t-j} + \sum_{k=1}^p \alpha_k Y_{t-k} + \gamma' X_t, \tag{9}$$

where F is a cumulative distribution function $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q \in \mathbb{R}$ and $\gamma \in \mathbb{R}^d$. We get the following result.

Proposition 2. *Assume that F takes values in $(0, 1)$, is positive everywhere and Lipschitz and that the covariate process X satisfies S1 and $\mathbb{E} \log_+ |X_1| < \infty$. Assume further that the roots of*

the polynomial

$$\mathcal{P}(z) = 1 - \sum_{j=1}^q \beta_j z^j$$

are outside the unit disc. There then exists a unique stationary solution to (9). Moreover, the process $((Y_t, X_t))_{t \in \mathbb{Z}}$ is stationary and ergodic.

Notes.

1. A classical choice for F is the Gaussian c.d.f. (probit model) or the logistic c.d.f. (logistic model). Model of type (9) have been proposed but without a theoretical support by Kauppi and Saikkonen [29], Rydberg and Shephard [42] or Russell and Engle [41] for analyzing price changes or predicting recessions. When $\beta_1 = \dots = \beta_q = 0$, a theory for the dynamic probit model can be found in de Jong and Woutersen [10] or in Fokianos and Truquet [22] who studied more general Markov models specified conditionally to some covariates. When there is no covariates, stationarity conditions for model (9) are given in Fokianos and Truquet [22]. Our results then extend these previous contributions and also give a theoretical basis to some models used in econometrics.
2. It is also possible to consider models with interactions between past values of the response and covariates. However, in general, application of Theorem 1 requires boundedness of the process $(\mu_t)_{t \in \mathbb{Z}}$ in (9). For simplicity, let us assume that $d = 1$ and that the process is (conditionally to X) a first-order time-inhomogeneous Markov chain (called a Markov chain with covariates in what follows) with $\mu_t = g(Y_{t-1}, X_t)$, $g : E \times \mathbb{R} \rightarrow \mathbb{R}$ being a measurable function. Then

$$b_0 = \sup_{z \in \mathbb{R}} d_{TV}(q(\cdot|1, z), q(\cdot|0, z)) = \sup_{z \in \mathbb{R}} |F(g(1, z)) - F(g(0, z))|.$$

Assumption S2 is valid provided the second link function g is bounded. When g is not bounded, Assumption S2 is still valid when for any $z \in \mathbb{R}$, $g(0, z)$ and $g(1, z)$ have the same sign but this restriction seems to be quite artificial. More generally, if $\mu_t = g(Y_{t-1}, \dots, Y_{t-p}, X_t)$ with $g : E^p \times \mathbb{R}^d \rightarrow \mathbb{R}$ is measurable and bounded, Assumption S2 is satisfied with $b_0 < 1$ and $b_m = 0$ if $m \geq p$. We point out that these results are less sharp than that of Fokianos and Truquet [22], where existence and uniqueness of a stationary and ergodic solution for a Markov chain with covariates was obtained without this boundedness assumption. However, Theorem 1 is compatible with non Markov processes and then observation-driven models which are more difficult to study.

We still consider the binary case as in (9) with a nonlinear latent process μ_t and with one lag for simplicity. For a function $g : \mathbb{R} \mapsto \mathbb{R}$, we assume that

$$\mu_t = g(\mu_{t-1}) + \alpha Y_{t-1} + \gamma' X_t. \tag{10}$$

Such type of model has been proposed by Russell and Engle [41] for analyzing financial transactions prices setting $g(s) = \beta s - \alpha F(s)$. Note that with the last specification, if $|\beta| < 1$, λ_t writes as a linear combination of the martingale differences $Y_{t-j} - F(\lambda_{t-j})$, $j \geq 1$.

Proposition 3. Assume that F takes values in $(0, 1)$, is Lipschitz and g is Lipschitz with

$$|g(s) - g(s')| \leq \kappa |s - s'|, \quad (s, s') \in \mathbb{R}^2$$

for some $\kappa \in (0, 1)$. Assume further that the process X satisfies Assumption S1 and $\mathbb{E} \log_+ |X_0| < \infty$. There exists a unique stationary process $(Y_t)_{t \in \mathbb{Z}}$ satisfying (10). Moreover, the process $((X_t, Y_t))_{t \in \mathbb{Z}}$ is stationary and ergodic.

4.2. Multinomial logistic autoregressions

We now provide a multinomial extension of the previous model. We consider the case of a state space $E = \{0, \dots, N - 1\}$ for an integer $N \geq 2$. For $i = 1, \dots, N - 1$, assume that

$$\begin{aligned} \mathbb{P}(Y_t = i | Y_{t-1}^-, X_t^-) &= \frac{\exp(\mu_{i,t})}{1 + \sum_{j=1}^{N-1} \exp(\mu_{j,t})}, \\ \mu_t &= \sum_{j=1}^q B_j \mu_{t-j} + \sum_{\ell=1}^p A_\ell \bar{Y}_{t-i} + \Gamma X_t, \end{aligned} \tag{11}$$

the B_j 's and the A_ℓ 's being matrices of size $(N - 1) \times (N - 1)$, Γ a matrix of size $(N - 1) \times d$. Moreover, \bar{Y}_{t-i} takes the k th column of the identity matrix I_{N-1} if Y_{t-i} takes the value k . In what follows, we denote by $\det(B)$ the determinant of a square matrix B . We have the following result. A proof is given in the supplementary material.

Proposition 4. Assume that the covariate process X satisfies S1 and $\mathbb{E} \log_+ |X_1| < \infty$. Assume further that the roots of the polynomial

$$\mathcal{P}(z) = \det \left(I_{N-1} - \sum_{j=1}^q B_j z^j \right)$$

are outside the unit disc. There then exists a unique stationary solution to (11). Moreover, the process $((Y_t, X_t))_{t \in \mathbb{Z}}$ is stationary and ergodic.

Note. This type of multinomial model is considered in Russell and Engle [41]. Let us point out that as for the multinomial regression, a modality of reference is chosen, here 0. In practice, the choice of this modality is often arbitrary and it is an undesirable property to have a model depending on this choice. Non-sensitivity to this choice requires that the differences $\mu_{i,t} - \mu_{j,t}$ for $i \neq j$ can be obtained via a change in the parameters of the specification of $\mu_{i,t}$. This is the case if $B_j = \beta_j I_{N-1}$ in (11), a condition also leading to a more parsimonious model. As for the binary case, more complex models can be obtained by including some interactions between past values of the response and the covariates. As already mentioned for binary models, to check our assumptions, it is in general necessary to assume boundedness of this interaction and then boundedness of the process μ_t .

4.3. Discrete choice models

Here, we assume that $E = \{1, \dots, N\}$. We want to consider stationary solutions of

$$Y_t = (\mathbb{1}_{\mu_{i,t} + \varepsilon_{i,t} > 0})_{1 \leq i \leq N}, \quad \mu_t = \sum_{j=1}^q B_j \mu_{t-j} + \sum_{k=1}^p A_k Y_{t-k} + \Gamma X_t, \quad (12)$$

where Γ is a matrix of size $N \times d$ and $A_1, \dots, A_p, B_1, \dots, B_q$ are square matrices of size $N \times N$. Such model is proposed for instance in Manner et al. [33], Candelon et al. [4] or Nyberg [37] for application to financial crisis, business cycles or recession dynamics. We will show it is possible to construct stationary paths for the dynamic (12) when the two process X and ε are independent. More precisely, setting for some $c \in \mathbb{R}^N$ and $I \subset E$,

$$C_I(c) = \bigcap_{i \in I} \{\varepsilon_{i,0} > -c_i\} \cap \bigcap_{i \in E \setminus I} \{\varepsilon_{i,0} \leq -c_i\},$$

we consider stationary processes $(Y_t)_{t \in \mathbb{Z}}$ solution of

$$\mathbb{P}(Y_t = \mathbb{1}_I | Y_{t-1}^-, X) = \mu(C_I(\mu_t)), \quad \mu = \mathbb{P}_{\varepsilon_0}, \quad (13)$$

where $\mathbb{1}_I$ is a vector of \mathbb{R}^N with a coordinate i equal to 1 if $i \in I$ and 0 otherwise. The following result is proved in the supplementary material.

Proposition 5. *Assume that X satisfies A1 with $\mathbb{E} \log_+ |X_0| < \infty$, ε_0 have a distribution with a full support \mathbb{R}^N and a Lipschitz c.d.f. Assume further that the roots of the polynomial $\mathcal{P}(z) = \det(I_N - \sum_{j=1}^q B_j z^j)$ are outside the unit disc. There then exists a unique stationary solution to the recursive equations (12), (13). Moreover, the process $((Y_t, X_t))_{t \in \mathbb{Z}}$ is stationary and ergodic.*

5. Measures of stochastic dependence

We will now study some dependence properties for chains with complete connections and exogenous covariates. In particular, such properties will be valid for the examples introduced in the previous section. Many dependence coefficients have been introduced in the literature. See Dedecker et al. [11] for a survey. The notion of strong mixing is probably one of most used for statistical applications. Doukhan [17] is a classical reference on this topic. However, strong mixing conditions are not always easy to check for a bivariate process of type $V_t = (Y_t, X_t)$. Under our assumptions, it is possible to show that the conditional probabilities $Y|X = x$ satisfy ϕ -mixing conditions. See Fokianos and Truquet [22] for a discussion in the homogeneous case, the arguments are the same here. However, there is no straightforward link between conditional and unconditional mixing. The notion of conditional mixing is considered for instance in Rao [39]. Yuan and Lei [48] give some counterexamples showing that conditional mixing properties do not necessarily entail unconditional strong mixing properties. Moreover, there exist autoregressive processes $(X_t)_{t \in \mathbb{Z}}$ that do not satisfy any strong mixing conditions and alternative dependence

coefficients have been proposed in the literature such as the functional dependence of Wu [47], adapted to Bernoulli shifts or the τ -dependence coefficients introduced by Dedecker and Prieur [12] and generalized in Dedecker and Prieur [13]. The latter dependence condition can be used as an alternative to the usual mixing conditions, since the usual deviations inequalities and invariance principles are available for partial sums of τ -dependent sequences. See, for instance, Dedecker and Prieur [12] and Merlevède et al. [34]. In this section, we will use our results for bounding either the coefficients of absolute regularity or τ -dependence, depending on the assumption made on the covariate process. Since we already pointed out the difficulty of getting unconditional dependence properties from that of conditional distributions, we will assume existence of a particular coupling of the covariate process instead of a particular weak dependence condition.

5.1. Absolute regularity and τ -dependence coefficients

For a stationary process $(V_t)_{t \in \mathbb{Z}}$ taking values in $E \times \mathbb{R}^d$ and $n \in \mathbb{N}^*$, we set

$$\beta_V(n) = \mathbb{E} \left[\sup_A |\mathbb{P}((V_n, V_{n+1}, \dots) \in A | \mathcal{F}_0) - \mathbb{P}((V_n, V_{n+1}, \dots) \in A)| \right],$$

where $\mathcal{F}_0 = \sigma(V_i : i \leq 0)$. We say that $(V_t)_{t \in \mathbb{Z}}$ is absolutely regular or β -mixing if $\lim_{n \rightarrow \infty} \beta_V(n) = 0$.

Next we recall the definition of the coefficients of τ -dependence. On $\bar{E} = E \times \mathbb{R}^d$, we consider the distance $\bar{\gamma}$ defined by

$$\bar{\gamma}(v, v') = \mathbb{1}_{v_1 \neq v'_1} + |v'_2 - v_2|.$$

We also define the following set of Lipschitz functions:

$$\mathcal{L}_\ell = \left\{ f : \bar{E}^\ell \rightarrow \mathbb{R} \text{ s.t. } \text{Lip}(f) := \sup_{w \neq w' \in \bar{E}^\ell} \frac{|f(w_1, \dots, w_\ell) - f(w'_1, \dots, w'_\ell)|}{\sum_{i=1}^\ell \bar{\gamma}(w_i, w'_i)} < \infty \right\}.$$

Finally, for a point $v_0 \in \bar{E}$, we set

$$\mathcal{P}_{0,\ell} = \left\{ \mu \text{ probability measure on } \bar{E}^\ell : \int \sum_{j=1}^\ell \bar{\gamma}(v_j, v_0) \mu(dv_1, \dots, dv_\ell) < \infty \right\}.$$

Note that the set $\mathcal{P}_{0,\ell}$ does not depend on the point v_0 . Next, for $\mu, \nu \in \mathcal{P}_{0,\ell}$, we define

$$\mathcal{W}_{1,\ell}(\mu, \nu) = \sup_{f \in \mathcal{L}_\ell} \left\{ \int f d\mu - \int f d\nu : \text{Lip}(f) \leq 1 \right\}.$$

Remember that from Kantorovich's duality, we have

$$\mathcal{W}_{1,\ell}(\mu, \nu) = \inf \left\{ \int \sum_{j=1}^\ell \bar{\gamma}(v_j, v'_j) \Gamma(dv_1, \dots, dv_\ell, dv'_1, \dots, dv'_\ell) \right\}$$

where the infimum is taken on the set of probability measures Γ on $\bar{E}^\ell \times \bar{E}^\ell$ having marginals μ and ν .

For $t \in \mathbb{Z}$, let $V_t = (Y_t, X_t)$ and $\mathcal{F}_t = \sigma(V_j : j \leq t)$. For an integer $\ell \geq 1$ and $j_1 < \dots < j_\ell$ in \mathbb{Z} , set $J = \{j_1, \dots, j_\ell\}$ and

$$U_J = (V_{j_1}, V_{j_2}, \dots, V_{j_\ell}).$$

According to Dedecker and Prieur [13], we define τ -dependence coefficients between U_J and \mathcal{F}_0 by

$$\tau(\mathcal{F}_0, U_J) = \mathbb{E}W_{1,\ell}(\mathbb{P}_{U_J|\mathcal{F}_0}, \mathbb{P}_{U_J}),$$

where $\mathbb{P}_{U_J|\mathcal{F}_0}$ denotes the conditional distribution of U_J given \mathcal{F}_0 . We then define for some integers $k \geq 1$ and $n \geq 1$,

$$\tau^{(k)}(n) = \max_{1 \leq \ell \leq k} \frac{1}{\ell} \sup \{ \tau(\mathcal{F}_0, U_{j_1, \dots, j_\ell}), n \leq j_1 < \dots < j_\ell \},$$

and $\tau_V(n) = \sup_{k \geq 1} \tau^{(k)}(n)$.

Note that the initial definition of the τ -dependence coefficients defined in Dedecker and Prieur [12] were defined when the distance $\bar{\gamma}$ is the ℓ_1 -metric. However, as E is a finite set, the two metrics are equivalent. Indeed, one can always code the elements of the finite set E as vectors of the canonical basis of \mathbb{R}^N (N is the number of elements of E) and in this case, we have, for $x, y \in E$,

$$\frac{1}{2} \sum_{i=1}^N |x_i - y_i| = \mathbb{1}_{x \neq y}.$$

Then one can assume that the process $(V_t)_{t \in \mathbb{Z}}$ takes values in \mathbb{R}^{N+d} and choose $\bar{\gamma}$ as the corresponding ℓ^1 -metric. In this case, Dedecker and Prieur [12] developed various limiting theorem for partial sums, when $\tau_V(n) \rightarrow 0$. Let us also mention that such dependence coefficients were generalized in Dedecker and Prieur [13], when the state space $(\bar{E}, \bar{\gamma})$ is a general Polish space.

5.2. Control of dependence coefficients

For bounding the dependence coefficients defined in the previous section, we will assume a representation of the form $X_t = g(S_t)$, where $(S_t)_{t \in \mathbb{Z}}$ is a stationary Markov chain taking values in a Polish space \mathcal{S} and $g : \mathcal{S} \rightarrow \mathbb{R}^d$ is a measurable function. We will denote by γ a metric on \mathbb{R}^d that will be either the discrete metric, that is, $\gamma(x_1, x_2) = \mathbb{1}_{x_1 \neq x_2}$, or the ℓ_1 -metric, that is, $\gamma(x_1, x_2) = |x_2 - x_1| := \sum_{i=1}^d |x_{i,1} - x_{i,2}|$ for $x_j = (x_{1,j}, \dots, x_{d,j}) \in \mathbb{R}^d$, $j = 1, 2$.

We introduce the probability kernel P from $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ to $(\mathcal{S}^{\mathbb{N}^*}, \mathcal{B}(\mathcal{S}^{\mathbb{N}^*}))$ and such for $s_0 \in \mathcal{S}$, $P(s_0, \cdot)$ is the distribution of $(S_t)_{t \geq 1}$ conditional on $S_0 = s_0$. Let also π be the invariant probability of the chain. We set $\Omega^* = \mathcal{S}^{\mathbb{N}^*} \times \mathcal{S}^{\mathbb{N}^*}$. For $t \in \mathbb{N}^*$, we denote by $S_{1,t} : \Omega \rightarrow \mathcal{S}$ and $S_{2,t} : \Omega^* \rightarrow \mathcal{S}$ the coordinate applications $S_{1,t}((s, \bar{s})) = s_t$ and $S_{2,t}((s, \bar{s})) = \bar{s}_t$.

S1'. We assume that there exists a probability kernel \tilde{P} from $(\mathcal{S}^2, \mathcal{B}(\mathcal{S}^2))$ to $(\Omega^*, \mathcal{B}(\Omega^*))$ such that $\tilde{P}((s_0, \bar{s}_0), \cdot)$ is a coupling of $P(s_0, \cdot)$ and $P(\bar{s}_0, \cdot)$ and

$$a_t := \int d\pi(s_0) d\pi(\bar{s}_0) \tilde{E}_{s_0, \bar{s}_0} [\gamma(g(S_{1,t}), g(S_{2,t}))]$$

satisfies $\lim_{t \rightarrow \infty} a_t = 0$. $\tilde{E}_{s_0, \bar{s}_0}$ denotes the expectation under $\tilde{P}((s_0, \bar{s}_0), \cdot)$.

S3. There exists a sequence $(e_m)_{m \in \mathbb{N}}$ such that $\sum_{m \geq 1} e_m < \infty$ and for all $(y, z, z') \in G^{\mathbb{N}} \times E^{\mathbb{N}} \times E^{\mathbb{N}}$,

$$d_{TV}(q(\cdot|y, z), q(\cdot|y, z')) \leq \sum_{i \geq 0} e_i |z_i - z'_i|.$$

Notes.

1. Assumption **S1'** guarantees the existence of a coupling of the chain with two independent initial states and for which the the distance between the future states goes to zero on average at infinity.
2. Assume that γ is induced by the ℓ_1 -norm on \mathbb{R}^d and the covariate process X is a Bernoulli shift. We remind that a process X is called a Bernoulli shift if there exists a measurable space Λ , a random sequence $\varepsilon \in \Lambda^{\mathbb{Z}}$ of i.i.d. random variables and a measurable mapping $g : \Lambda^{\mathbb{N}} \rightarrow G$ such that

$$X_t = g(\varepsilon_t, \varepsilon_{t-1}, \dots), \quad t \in \mathbb{Z}.$$

We point out that such a representation is valid for many time series models found in the literature from linear processes of ARMA type to GARCH processes. In this case, we set $S_t = (\varepsilon_t, \varepsilon_{t-1}, \dots)$ which takes values in $\mathcal{S} = \Lambda^{\mathbb{N}}$. Here, the kernel \tilde{P} in Assumption **S1'** can be defined as the probability distribution of $(\varepsilon^{(s_0)}, \varepsilon^{(\bar{s}_0)})$ where for $s \in \Lambda^{\mathbb{N}}$, $\varepsilon_t^{(s)} = \varepsilon_t$ for $t \geq 1$ and $\varepsilon_t^{(s)} = s_{-t}$ for $t \leq 0$. In this case, we have the expression $a_t = \mathbb{E}(|X_t - \bar{X}_t|)$ where $t \geq 1$,

$$\bar{X}_t = g(\varepsilon_t, \dots, \varepsilon_1, \varepsilon'_0, \varepsilon'_{-1}, \dots)$$

and ε' is an independent copy of ε . A martingale argument shows that the condition $\lim_{t \rightarrow \infty} a_t = 0$ is automatically satisfied.

3. When $\gamma(x_1, x_2) = \mathbb{1}_{x_1 \neq x_2}$, Assumption **S1'** is implied by the existence of a so-called successful coupling of two chains with different initial values. See Lindvall [32] for a discussion of existence of successful coupling for Markov chains and in particular Theorem 14.10 which shows an equivalence with the weak ergodicity property of the Markov chain. Existence of a successful coupling means that there exists a probability kernel \tilde{P} , defined as in Assumption **S1'** and such that

$$\tilde{P}((s_0, \bar{s}_0), \{\exists n_0 \in \mathbb{N} : S_{1,n} = S_{2,n}, n \geq n_0\}) = 1.$$

Then the condition $\lim_{t \rightarrow \infty} a_t = 0$ is automatically satisfied for such kernel. Indeed, we have

$$\tilde{E}_{s_0, \bar{s}_0}[\gamma(g(S_{1,t}), g(S_{2,t}))] \leq \tilde{P}\left((s_0, \bar{s}_0), \bigcup_{i \geq t} \{S_{1,i} \neq S_{2,i}\}\right) \rightarrow 0.$$

and we get $a_t \rightarrow 0$ from Lebesgue's theorem.

4. One can also assume that $S_t = (\varepsilon_t, \varepsilon_{t-1}, \dots)$ where $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a homogenous chain with complete connections satisfying Assumption S2. Using Lemma 1, one can check Assumption S1' when $g(S_t) = h(\varepsilon_t, \dots, \varepsilon_{t-k})$ for some integer k and function h .

In what follows, we denote by \mathbb{L}^p the Lebesgue space of random variables taking values in \mathbb{R}^d and possessing a moment of order p (or bounded a.s. if $p = \infty$) and $\|\cdot\|_p$ the corresponding norm.

Theorem 2. *Assume that Assumptions S1–S1' and S2–S3 hold true.*

1. *Assume that γ is the discrete metric and that $X_0 \in \mathbb{L}^p$ for some $p \in [1, \infty]$. Set $q = \frac{p}{p-1}$ and $c_t = \max(1, 2\|X_0\|_p)a_t^{1/q}$. Then, for $n \in \mathbb{N}^*$, we have*

$$\beta_V(n) \leq \sum_{j \geq n} g_j,$$

with $g_j = b_{j-1}^* + c_j + \kappa_j + \sum_{i=0}^{j-2} b_i^* \kappa_{j-i-1}$ and $\kappa_j = \sum_{s=0}^{j-1} e_s c_{j-s} + 2 \sum_{s \geq j} e_s \mathbb{E}[|X_0|]$.

2. *Assume that γ is induced by the ℓ_1 -norm. Then, for $n \in \mathbb{N}^*$, we have*

$$\tau_V(n) \leq \sup_{j \geq n} h_j,$$

with $h_j = b_{j-1}^* + a_j + \kappa_j + \sum_{i=0}^{j-2} b_i^* \kappa_{j-i-1}$, $\kappa_j = \sum_{s=0}^{j-1} e_s a_{j-s} + 2 \sum_{s \geq j} e_s \mathbb{E}[|X_0|]$.

Notes.

1. Note that under our assumptions, $\lim_{i \rightarrow \infty} h_i = 0$ and then $\lim_{n \rightarrow \infty} \tau_V(n) = 0$. This is essentially due to the fact that if $(u_n)_{n \geq 0}$ is a summable sequence of nonnegative real numbers and $(v_n)_{n \in \mathbb{N}}$ a sequence of nonnegative real numbers converging to 0, then $\lim_{n \rightarrow \infty} \sum_{i=0}^n u_i v_{n-i} = 0$.
2. Under the assumptions of Theorem 2, the process V is absolutely regular provided that $\sum_{j \geq 1} g_j < \infty$. This is the case if $\sum_{j \geq 1} j e_j < \infty$ and $\sum_{j \geq 0} c_j < \infty$.
3. To get a precise rate of convergence for the dependence coefficients, it is necessary to control the decay of b_m^* when $m \rightarrow \infty$. Lemma 4 given in the supplementary material provides some results. In particular, when $b_m = O(m^{-k})$ for some $k \in \mathbb{N}^*$, we also have $b_m^* = O(m^{-k})$.

Next we give a result focused on observation-driven models. Our conditions will be specified for the examples already mentioned in Section 4

Corollary 2. Assume that Assumptions S1–S1' hold true. Suppose that $q(\cdot|Y_{t-1}^-, X_t^-) = \check{q}(\cdot|\lambda_t)$ is Lipschitz in λ_t with $\lambda_t = G_{Y_{t-1}, X_t}(\lambda_{t-1})$ and the assumptions of Lemma 2 are satisfied.

1. Assume that S1' holds true for the discrete metric. If $a_i = O(i^{-\kappa})$ with $\kappa > q$, $X_0 \in \mathbb{L}_p$ and $p^{-1} + q^{-1} = 1$, we have $\beta_V(n) = O(n^{-\kappa/q+1})$ and the process V is absolutely regular. If $a_i = O(\rho^i)$ for some $\rho \in (0, 1)$, then $\beta_V(n) = O(\bar{\rho}^n)$ for some $\bar{\rho} \in (0, 1)$.
2. Assume that S1' holds true for the ℓ_1 -metric. If $a_i = O(i^{-\kappa})$ with $\kappa > 1$, we have $\tau_V(n) = O(n^{-\kappa})$. If $a_i = O(\rho^i)$ for some $\rho \in (0, 1)$, then $\tau_V(n) = O(\bar{\rho}^n)$ for some $\bar{\rho} \in (0, 1)$.

Note. These dependence coefficients are useful to derive limit theorems or deviation inequalities for some partial sums of the form

$$S_n = \frac{1}{n} \sum_{t=1}^n W_t, \quad W_t := f(V_t, V_{t-1}, \dots), \quad V_t = (Y_t, X_t)$$

for some suitable real-valued functions f , such as some smooth functionals of $(Y_{t-j}, \mu_{t-j})_{j \geq 0}$, where μ denotes the latent process of observation-driven models satisfying our assumptions. We point out that, due to the discrete nature of the process $(Y_t)_{t \in \mathbb{Z}}$, the process $(W_t)_{t \in \mathbb{Z}}$ is not necessarily absolutely regular. Neumann [36] studied this problem for the latent process of a Poisson autoregressive process. However, limit theorems for S_n can be still be obtained in approximating W_t by some functionals only depending on finitely many V_{t-j} , $j \geq 0$. On the other hand, it is still possible to control directly the τ -dependence coefficients of $(W_t)_t$ when the function f is smooth enough. See Section 2 in the supplementary material.

6. Perspectives in statistics

Let us now give some possible applications of our results to statistical inference in the models presented in Section 4.

1. The first problem concerns parametric estimation which has been extensively studied for other observation-driven models. Usually, only ergodicity is necessary to get consistency and asymptotic normality of the conditional likelihood estimator. For instance, Douc et al. [16] studied this problem when some general observation-driven models are well specified and misspecified. Then one could obtain similar results for our models satisfying the Assumptions S1–S2. Let us also mention that models of infinite order can be considered, such as (8), with a parametric form for the parameters $a_j = a_j(\theta)$ and a decay in the dependence which is not exponential (in contrast to observation-driven models). For instance, $a_j(\theta) = \theta_1 j^{-\theta_2}$ with $\theta_2 > 2$.
2. The second problem concerns discrete choice models as in (12) and the estimation of the distribution of ε_0 modeled via a parametric copula, as in Manner et al. [33]. For instance, the distribution of ε_t can be marginally Gaussian or logistic and we obtain a multivariate version of the univariate probit or logistic binary model. Our results can then be used to study the asymptotic properties of the maximum likelihood estimator given in their paper.

3. The third problem concerns semi-parametric estimation in our models. Recently, Park et al. [38] investigated a related nonparametric problem but for finite-order models. Let us describe an approach for the binary time series models given in Section 4. Our aim is to estimate the function F as well as a vector θ of autoregressive parameters. One can then maximize

$$\theta \mapsto \sum_{t=1}^n [Y_t \log \hat{F}_\theta(\mu_t(\theta)) + (1 - Y_t) \log(1 - \hat{F}_\theta(\mu_t(\theta)))]$$

with

$$\hat{F}_\theta(z) = \frac{\sum_{t=1}^n Y_t K_h(z - \mu_t(\theta))}{\sum_{t=1}^n K_h(z - \mu_t(\theta))}$$

and K is a kernel, $h > 0$ a bandwidth parameter and $K_h = h^{-1}K(\cdot/h)$. If $\hat{\theta}$ is such maximizer, $\mathbb{P}(Y_t = 1 | Y_{t-1}^-, X_t^-)$ can be estimated by $\hat{F}_{\hat{\theta}}(\mu_t(\hat{\theta}))$. The dependence properties stated in Section 5 will be essential to derive asymptotic properties of this estimator. With respect to the problem considered in Park et al. [38], the main interest of this semi-parametric approach is that one can use much more lags values for the response and the covariates in order to predict the Y_t 's.

7. Proofs of the results

7.1. Proof of Lemma 1

Without loss of generality, we will assume that $t_0 = 0$, the general case will follow by replacing t by $t - t_0$ in the bound we will derive. We then remove the index t_0 from all our notations. We will apply the technique of maximal coupling already used by Bressaud et al. [1] for getting a bound on the relaxation speed of chains with complete connections. We defer the reader to Bressaud et al. [1], Equation 4.9 for a precise definition of the maximal coupling of two probability measures α and $\bar{\alpha}$ on the finite set E . In what follows, we will simply use the fact that there exists a probability measure $\alpha \tilde{\times} \bar{\alpha}$ on $E \times E$ such that

$$\alpha \tilde{\times} \bar{\alpha}(\{(y, \bar{y}) \in E^2 : y \neq \bar{y}\}) = d_{TV}(\alpha, \bar{\alpha}).$$

For a sequence $\omega = (\omega_{i,j})_{i,j \geq 0} \in E^{\mathbb{N} \times \mathbb{N}^*}$, we denote, for $(j, k) \in \mathbb{N} \times \mathbb{N}^*$, by $\omega_{j,1:k}$ the vector $(\omega_{j,1}, \dots, \omega_{j,k}) \in E^k$. We then set

$$\Gamma_{0,1}^{(k)}(\omega_{0,1:k}) = \prod_{t=1}^k q_t^x(\omega_{0,t} | \omega_{0,t-1}^-).$$

In the previous expressions and the next ones, we always use the convention $\omega_{0,-i} = z_i$ and $\omega_{j,-i} = \bar{z}_i$ for $i \geq 0$ and $j \geq 1$. $\Gamma_{0,1}^{(k)}$ defines a probability measure on E^k . Next, we define k

probability kernels $\Gamma_1^{(k)}, \dots, \Gamma_k^{(k)}$ on E^k in the following way.

$$\Gamma_1^{(k)}(\omega_{1,1:k}|\omega_{0,1:k}) = \frac{\prod_{t=1}^k [q_t^x(\cdot|\omega_{0,t-1}^-) \tilde{\times} q_t^x(\cdot|\omega_{1,t-1}^-)](\omega_{0,t}, \omega_{1,t})}{\Gamma_{0,1}^{(k)}(\omega_{0,1:k})}.$$

If $2 \leq j \leq k$, the kernel $\Gamma_j^{(k)}$ is defined by the equality

$$\begin{aligned} \Gamma_j^{(k)}(\omega_{j+1,1:k}|\omega_{j,1:k}) &\times \prod_{t=1}^{j-1} \bar{q}_t^x(\omega_{j,t}|\omega_{j,t-1}^-) \cdot \prod_{t=j}^k q_t^x(\omega_{j,t}|\omega_{j,t-1}^-) \\ &= \prod_{t=1}^{j-1} [\bar{q}_t^x(\cdot|\omega_{j,t-1}^-) \tilde{\times} \bar{q}_t^x(\cdot|\omega_{j+1,t-1}^-)](\omega_{j,t}, \omega_{j+1,t}) \\ &\quad \times [q_j^x(\cdot|\omega_{j,j-1}^-) \tilde{\times} \bar{q}_j^x(\cdot|\omega_{j+1,j-1}^-)](\omega_{j,j}, \omega_{j+1,j}) \\ &\quad \times \prod_{t=j+1}^k [q_t^x(\cdot|\omega_{j,t-1}^-) \tilde{\times} q_t^x(\cdot|\omega_{j+1,t-1}^-)](\omega_{j,t}, \omega_{j+1,t}). \end{aligned}$$

Finally, we define a probability measure $P_{x,z,\bar{z}}^{(k)}$ on $(E^k)^{k+1}$ by

$$P_{x,z,\bar{z}}^{(k)}(\omega_{0,1:k}, \dots, \omega_{k+1,1:k}) = \Gamma_{0,1}^{(k)}(\omega_{0,1:k}) \times \prod_{j=0}^k \Gamma_j^{(k)}(\omega_{j+1,1:k}|\omega_{j,1:k}).$$

Let us give an interpretation of the measure $P_{x,z,\bar{z}}^{(k)}$. This measure is the probability distribution of the $k + 1$ first coordinates of a Markov chain on the state space E^k . Each coordinate of the chain can be seen as a path of a chain with complete connection.

- $\Gamma_{0,1}^{(k)}$ is the distribution of k successive coordinates of a chain with complete connection with initialization z_{-i} for $i \leq 0$ and transition kernels q_1^x, \dots, q_k^x .
- The joint distribution of the first path and the second path is obtained by applying iteratively the maximal coupling to the transition kernels (q_t^x, \bar{q}_t^x) from time $t = 1$ to time $t = k$. The second path is initialized with \bar{z}_{-i} for $i \leq 0$. The second path has then the same transition kernels as the first path but a different initialization.
- For $1 \leq j \leq k$, the j th path is initialized with \bar{z}_{-i} , $i \leq 0$ and has transition kernels $\bar{q}_1^x, \dots, \bar{q}_{j-1}^x, q_j^x, \dots, q_k^x$. The path $j + 1$ is obtained as the path j , except that at time $t = j$, the kernel q_t^x is replaced with the kernel \bar{q}_t^x . The joint probability distribution of the paths j and $j + 1$ is obtained by applying iteratively the maximal coupling to these transition kernels.

Our approach is equivalent to make several couplings of two successive paths having either a different initialization or one transition kernel changing across the time and then “gluing” all the paths to define a joint probability distribution on $(E^k)^{k+1}$. Our definition of this joint probability

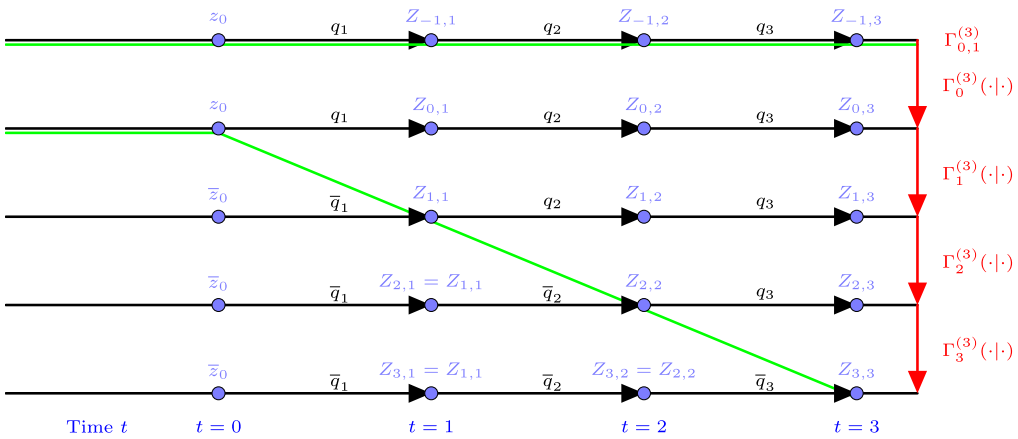


Figure 1. Description of the coupling when $k = 3$. Γ_i , which denotes the coupling between the path $i - 1$ and the path i , is obtained by applying iteratively the maximal coupling over the time index t . For $i = 0$, only the distribution of past values ($t \leq 0$) is changed ($(Z_{-1,j})_{j \leq 0}$ and $(Z_{0,j})_{j \leq 0}$ are assumed to be independent) whereas for $i = 1, 2, 3$, only one conditional distribution is modified between two successive paths. The gluing technique allows to define all the paths on the same probability space by using the conditional distribution of the coupling measures. The two green lines correspond to the two paths (y_t) and (\bar{y}_t) .

measure is classical in coupling theory and can be seen as a particular application of the so-called gluing lemma. See Villani [45], Chapter 1. It is much easier to visualize such a coupling graphically. Figure 1 gives a description of this coupling scheme when $k = 3$.

Next, let us observe that $(x, z, \bar{z}) \mapsto P_{x,z,\bar{z}}^{(k)}$ is measurable. This is a consequence of the definition of $P_{x,z,\bar{z}}^{(k)}$ and of the explicit expression of the maximal coupling of two discrete probability measures in term of the marginals. Measurability of the previous mapping then follows from Assumption A1. We now mention that the sequence $(P_{x,z,\bar{z}}^{(k)})_{k \geq 1}$ satisfies Kolmogorov’s compatibility conditions. Indeed, one can show that

$$P_{x,z,\bar{z}}^{(k)}(\omega_{0,1:k}, \dots, \omega_{k+1,1:k}) = \sum_{\omega_{k+2,1:k+1} \in E^{k+1}} \sum_{\omega_{0,k+1,k+1} \in E^{k+2}} P_{x,z,\bar{z}}^{(k+1)}(\omega_{0,1:k+1}, \dots, \omega_{k+2,1:k+1}).$$

From the Kolmogorov extension theorem, there exists a unique probability measure $P_{x,z,\bar{z}}$ on $E^{\mathbb{N} \times \mathbb{N}^*}$ compatible with this sequence. Note that, for any $A \in \mathcal{P}(E)^{\mathbb{N} \times \mathbb{N}^*}$, the mapping $(x, z, \bar{z}) \mapsto P_{x,z,\bar{z}}(A)$ is still measurable. This was already justified when A is a cylinder set. Extension of the measurability for A arbitrary follows from a monotone class argument.

Now for $\omega \in E^{\mathbb{N} \times \mathbb{N}^*}$ and $j \geq -1, k \geq 1$, we define $Z_{j,k}(\omega) = \omega_{j+1,k}$. We define the probability distribution $\tilde{Q}_{x,z,\bar{z}}$ as the pushforward measure of $P_{x,z,\bar{z}}$ obtained from $((Z_{-1,t})_{t \geq 1}, (Z_{t,t})_{t \geq 1})$. Note that, from our construction with the maximal coupling, we have automatically $Z_{k,t} = Z_{t,t}$ for $k \geq t \geq 0$, $P_{x,z,\bar{z}}$ a.s. Indeed, when the two past sequences are equal, the maximal coupling generates two identical random variables. We then deduce that $(Z_{t,t})_{t \geq 1}$ has transition

kernels $(\bar{q}_t^x)_{t \geq 1}$. This proves (4). Let us now prove the bound (5). Let t be a positive integer. We denote by $E_{x,z,\bar{z}}$ the mathematical expectation under $P_{x,z,\bar{z}}$. From the triangular inequality, we have

$$\tilde{Q}_{x,z,\bar{z}}(\{y_t \neq \bar{y}_t\}) = P_{x,z,\bar{z}}(Z_{-1,t} \neq Z_{t,t}) \leq \sum_{k=-1}^{t-1} P_{x,z,\bar{z}}(Z_{k,t} \neq Z_{k+1,t}). \tag{14}$$

If $k = -1$, one can use Lemma 3 given in the supplementary material to get

$$P_{x,z,\bar{z}}(Z_{-1,t} \neq Z_{0,t}) \leq b_{t-1}^*.$$

Indeed $Z_{-1,\cdot}$ and $Z_{0,\cdot}$ are constructed using the maximal coupling and when $t \geq 1$, the transition kernel for the two paths equals $q_t^x(\cdot|\cdot)$. Lemma 3 is in fact a variation of a result given in Bressaud et al. [1] for controlling the relaxation speed of a homogenous chain with complete connections. A detailed proof of this lemma is given in the supplementary material. If $t = k + 1 \geq 1$, we have from the definition of the maximal coupling and from our construction

$$\begin{aligned} P_{x,z,\bar{z}}(Z_{k,t} \neq Z_{k+1,t}) &= E_{x,z,\bar{z}}[P_{x,z,\bar{z}}(Z_{k,t} \neq Z_{k+1,t} | \sigma(Z_{k,t-j}, Z_{k+1,t-j}; j \geq 1))] \\ &\leq \sup_{w \in E^{\mathbb{N}}} d_{\text{TV}}(q_t^x(\cdot|w), \bar{q}_t^x(\cdot|w)). \end{aligned}$$

Next, if $t \geq k + 2 \geq 2$, we have

$$\begin{aligned} P_{x,z,\bar{z}}(Z_{k,t} \neq Z_{k+1,t}) &= E_{x,z,\bar{z}}[P_{x,z,\bar{z}}(Z_{k,t} \neq Z_{k+1,t} | \sigma(Z_{k,k+1-j}, Z_{k+1,k+1-j}; j \geq 0)) \mathbb{1}_{Z_{k,k+1} \neq Z_{k+1,k+1}}] \\ &\leq b_{t-k-2}^* P_{x,z,\bar{z}}(Z_{k,k+1} \neq Z_{k+1,k+1}) \\ &\leq b_{t-k-2}^* \sup_{w \in E^{\mathbb{N}}} d_{\text{TV}}(q_{k+1}^x(\cdot|w), \bar{q}_{k+1}^x(\cdot|w)). \end{aligned}$$

Let us comment the previous bounds. The first equality follows from the fact that on the event $\{Z_{k,k+1} = Z_{k+1,k+1}\}$, we automatically have $Z_{k,j} = Z_{k+1,j}$ for $j \geq k + 1$. This is due to the maximal coupling and to the fact that, from our construction, we have $Z_{k,j} = Z_{k+1,j}$ for $j \leq k$. The second bound follows from Lemma 3 and the fact that for $s = k + 2, \dots, t$, the two paths $Z_{k,\cdot}$ and $Z_{k+1,\cdot}$ have the same transition kernels $q_s^x(\cdot|\cdot)$, $s \geq 1$. Finally, the third bound follows from the definition of the maximal coupling. The bound (5) follows from (14) and our previous bounds. Finally, the third point of the lemma follows from the measurability properties of $(x, z, \bar{z}) \mapsto P_{x,z,\bar{z}}$. The proof of Lemma 1 is now complete.

7.2. Proof of Theorem 1

1. Let $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ be a probability space on which the covariate process X is defined. For simplicity, we assume that $\bar{\Omega} = G^{\mathbb{Z}}$ is the canonical space of the paths. We then have $X_t(\omega) = \omega_t$ for all $(t, \omega) \in \mathbb{Z} \times \bar{\Omega}$. Existence of a stochastic process $(Y_t)_{t \in \mathbb{Z}}$ satisfying (7) is understood as follows. We consider an enlargement $(\Omega, \mathcal{A}, \mathbb{P})$ of the initial probability

space with $\Omega = E^{\mathbb{Z}} \times \overline{\Omega}$, $\mathcal{A} = \mathcal{P}(E)^{\otimes \mathbb{Z}} \otimes \overline{\mathcal{A}}$. For all $(y, \overline{\omega}, t) \in \Omega \times \mathbb{Z}$, we set $Y_t(y, \overline{\omega}) = y$, $X'_t(y, \overline{\omega}) = X_t(\overline{\omega})$ and \mathbb{P} is the probability measure such that $\mathbb{P}(X' \in E^{\mathbb{Z}} \times B) = \overline{\mathbb{P}}(X \in B)$ for all $B \in \overline{\mathcal{A}}$ and for which (7) is satisfied, replacing X with X' . For getting uniqueness, we prove the following property. If $(\Omega', \mathcal{A}', \mathbb{P}')$ is another probability space on which two processes X' and Y are defined and satisfy $\mathbb{P}'(X' \in B) = \mathbb{P}(X \in B)$ and (7), then $\mathbb{P}'((Y', X') \in A) = \mathbb{P}((Y, X) \in A)$ for all $A \in \mathcal{P}(E)^{\otimes \mathbb{Z}} \otimes \mathcal{B}(G)^{\otimes \mathbb{Z}}$.

To show these properties, we will construct, for each $x \in \mathcal{D}$, a family of finite-dimensional probability distributions $\{v_x^I : I = \{s+1, \dots, s+n\}; (s, n) \in \mathbb{Z} \times \mathbb{N}^*\}$ such that for all I , v_x^I is a probability measure on E^I and satisfies the compatibility conditions of the Kolmogorov's extension theorem. To this end, we will use the coupling result of Lemma 1 and more precisely the control in total variation given in (6). Let us consider a set $I = \{s+1, \dots, s+n\}$ of n successive integers and two elements z, \overline{z} of $E^{\mathbb{N}}$. For an integer $i \geq 2$, set $t_0 = s - i$ and for $t \geq t_0 + 1$ and $y \in E^{\mathbb{Z}}$,

$$q_t^x(y_0|y_{-1}^-) = \overline{q}_t^x(y_0|y_{-1}^-) = q(y_0|y_{-1}^-, x_t^-(\omega)).$$

We define a probability measure $v_{x,i,z}^I$ by

$$v_{x,i,z}^I(y_{s+1}, \dots, y_{s+n}) = \int \sum_{y_{t_0+1}, \dots, y_s} \prod_{t=t_0+1}^{s+n} q_t^x(y_t|y_{t-1}^-) \delta_z(dy_{t_0}^-).$$

Using (6), we have

$$d_{\text{TV}}(v_{x,i,z}^I, v_{x,i,\overline{z}}^I) \leq \sum_{\ell=0}^{n-1} b_{i+\ell}^*.$$

Assumption S2 guarantees the summability of the b_m^* 's and hence that $b_m^* \rightarrow 0$. Moreover, the previous bound does not depend on the couple (z, \overline{z}) and goes to zero when $i \rightarrow \infty$, one can show that the sequence $(v_{x,i,z}^I)_{i \geq 2}$ is a Cauchy sequence in the simplex of \mathbb{R}^{N^n} and has a limit which does not depend on z . We then set

$$v_x^I = \lim_{i \rightarrow \infty} v_{x,i,z}^I. \quad (15)$$

The compatibility conditions on the family of finite dimensional distributions

$$\mathcal{G} = \{v_x^I : I = \{s+1, \dots, s+n\}, (s, n) \in \mathbb{Z} \times \mathbb{N}^*\}$$

follows from the fact that for any $i \geq 2$ and $I = \{s+1, \dots, s+n\}$,

$$\begin{aligned} \sum_{y_{s+n+1} \in E} v_{x,z,i}^{I \cup \{s+n+1\}}(y_{s+1}, \dots, y_{s+n+1}) &= v_{x,z,i}^I(y_{s+1}, \dots, y_{s+n}), \\ \sum_{y_s \in E} v_{x,z,i}^{I \cup \{s\}}(y_s, \dots, y_{s+n}) &= v_{x,z,i+1}^I(y_{s+1}, \dots, y_{s+n}). \end{aligned}$$

The Kolmogorov’s extension theorem guarantees existence and unicity of a probability measure ν_x on $(\mathbb{E}^{\mathbb{Z}}, \mathcal{P}(E)^{\mathbb{Z}})$ compatible with the family \mathcal{G} . We then $\Omega = E^{\mathbb{Z}} \times \overline{\Omega}$ and for $\omega = (y, \overline{\omega}) \in \Omega$, $\mathbb{P}(d\omega) = \nu_{X(\overline{\omega})}(dy)\overline{\mathbb{P}}(d\overline{\omega})$ and $Y_t(\omega) = y_t$ for $t \in \mathbb{Z}$. We point out that measurability of the mapping $x \mapsto \nu_x(A)$ can be shown first when A is a cylinder set and then for an arbitrary $A \in \mathcal{P}(E)^{\otimes \mathbb{Z}}$ using a monotone class argument.

- Next, we show that the process $(Y_t)_{t \in \mathbb{Z}}$ defined in the previous point satisfies (7). This is equivalent to show that the probability measure ν_x defined in the previous point is compatible with the sequence $(q_t^x)_t$. We keep the notations of the previous point. Fix $k \in \mathbb{N}^*$. Let $\epsilon \in (0, 1)$ and $h : E \rightarrow \mathbb{R}_+$ and $g : E^k \rightarrow \mathbb{R}_+$ be some functions taking nonnegative values and bounded by one. Let \overline{y} be an arbitrary element of E . From our assumptions, there exists and integer $m > k$ such that $b_m \leq \epsilon$ if $m \geq k$. Set

$$q_{m,t}^x(y_t|y_{t-1:t-m}) = q_t^x(y_t|y_{t-1}, \dots, y_{t-m}, \overline{y}, \overline{y}, \dots).$$

We also set $I = \{t - k, \dots, t\}$, $I_m = \{t - m, \dots, t - 1\}$ and we choose $i > m - k$ large enough such that

$$d_{\text{TV}}(\nu_x^I, \nu_{x,z,i}^I) + d_{\text{TV}}(\nu_x^{I_m}, \nu_{x,z,k+i-m}^{I_m}) \leq \epsilon.$$

We have

$$\left| \int h(y_t)g(y_{t-1}, \dots, y_{t-k}) [d\nu_x^I(y_{t-k}, \dots, y_t) - d\nu_{x,z,i}^I(y_{t-k}, \dots, y_t)] \right| \leq \epsilon.$$

Next we set $q_{m,t}^x h(y_{t-1}, \dots, y_{t-m}) = \sum_{y_t \in E} h(y_t)q_{m,t}^x(y_t|y_{t-1:t-m})$ and

$$A_m = \int q_{m,t}^x h(y_{t-1}, \dots, y_{t-m})g(y_{t-1}, \dots, y_{t-k}) d\nu_{x,z,k+i-m}^{I_m}(y_{t-m}, \dots, y_{t-1}).$$

We have

$$\left| A_m - \int h(y_t)g(y_{t-1}, \dots, y_{t-k}) d\nu_{x,z,i}^I(y_{t-k}, \dots, y_t) \right| \leq b_m \leq \epsilon.$$

Moreover

$$\left| A_m - \int q_{m,t}^x h(y_{t-1}, \dots, y_{t-m})g(y_{t-1}, \dots, y_{t-k}) d\nu_x^{I_m}(y_{t-m}, \dots, y_{t-1}) \right| \leq \epsilon.$$

Using the fact that,

$$|q_{m,t}^x h(y_{t-1}, \dots, y_{t-m}) - q_t^x h(y_{t-1}^-)| \leq b_m \leq \epsilon,$$

we get

$$\left| \int h(y_t)g(y_{t-1}, \dots, y_{t-k}) d\nu_x^I(dy_t, \dots, dy_{t-k}) - \int q_t^x h(y_{t-1}^-)g(y_{t-1}, \dots, y_{t-k}) d\nu_x(y) \right| \leq 4\epsilon.$$

This proves that

$$\int h(y_t)g(y_{t-1}, \dots, y_{t-k}) dv_x^I(y_{t-k}, \dots, y_t) = \int q_t^x h(y_{t-1}^-)g(y_{t-1}, \dots, y_{t-k}) dv_x(y).$$

From a monotone class argument, we obtain (7).

3. The equality between the two conditional expectations in point 1 of Theorem 1 is a consequence of the expression of v_x and q_t^x (which only depends of x_t^-).
4. Let us now show that the process $(V_t)_{t \in \mathbb{Z}}$ defined by $V_t = (Y_t, X_t)$ is stationary. It should be noticed first that if $I_t = \{t + 1, \dots, t + n\}$ for $t \in \mathbb{Z}$, then from the definition of the finite-dimensional distributions, we have

$$v_x^{I_t} = v_{\tau^t x}^{I_0} \quad \text{a.s.}$$

where $\tau^t x = (x_{t+j})_{j \in \mathbb{Z}}$. We then get for a measurable and bounded function $h : E^n \times \mathcal{D} \rightarrow \mathbb{R}$,

$$\begin{aligned} \mathbb{E}h(Y_{t+1}, \dots, Y_{t+n}, \tau^t X) &= \sum_{y_1, \dots, y_n \in E} \int h(y_1, \dots, y_n, \tau^t X(\bar{\omega})) v_{X(\bar{\omega})}^{I_t}(y_1, \dots, y_n) d\bar{\mathbb{P}}(\bar{\omega}) \\ &= \sum_{y_1, \dots, y_n \in E} \int h(y_1, \dots, y_n, \tau^t X(\bar{\omega})) v_{\tau^t X(\bar{\omega})}^{I_0}(y_1, \dots, y_n) d\bar{\mathbb{P}}(\bar{\omega}) \\ &= \sum_{y_1, \dots, y_n} \int h(y_1, \dots, y_n, X(\bar{\omega})) v_{X(\bar{\omega})}^{I_0}(y_1, \dots, y_n) d\bar{\mathbb{P}}(\bar{\omega}) \\ &= \mathbb{E}h(Y_1, \dots, Y_n, X). \end{aligned}$$

This shows the stationarity of the process $(V_t)_{t \in \mathbb{Z}}$.

5. Next, let us show uniqueness. Let $((Y'_t, X'_t))_{t \in \mathbb{Z}}$ be a stationary process satisfying the same assumptions. Setting $z_i = Y'_{t-i}$, we know that from (15), we have a.s.,

$$\begin{aligned} \lim_{i \rightarrow \infty} \mathbb{P}'(Y'_{t+1} = y_1, \dots, Y'_{t+n} = y_n | \sigma(X', Y'_{t-j} : j \geq i)) &= \lim_{i \rightarrow \infty} v_{X', i, z_i}^{I_t}(y_1, \dots, y_n) \\ &= v_{X'}^{I_t}(y_1, \dots, y_n). \end{aligned}$$

Hence for any measurable and bounded function h , we have

$$\begin{aligned} \mathbb{E}'h(Y'_{t+1}, \dots, Y'_{t+n}, X') &= \lim_{i \rightarrow \infty} \mathbb{E}'[\mathbb{E}'(h(Y'_{t+1}, \dots, Y'_{t+n}, X') | \sigma(X', Y'_{t-j} : j \geq i))] \\ &= \sum_{y_1, \dots, y_n} \mathbb{E}'[v_{X'}^{I_t}(y_1, \dots, y_n) h(y_1, \dots, y_n, X')] \\ &= \sum_{y_1, \dots, y_n} \mathbb{E}[v_X^{I_t}(y_1, \dots, y_n) h(y_1, \dots, y_n, X)] \\ &= \mathbb{E}h(Y_{t+1}, \dots, Y_{t+n}, X). \end{aligned}$$

The second equality follows from the Lebesgue theorem.

6. For $t \in \mathbb{Z}$, we remind that $V_t = (Y_t, X_t)$. We now prove the ergodicity property for the process $(V_t)_{t \in \mathbb{Z}}$. To this end, we adapt the direct proof of Kifer [30] who proved ergodic properties of some Markov chains in random environments. Set $\mu = \mathbb{P}_X$, the probability distribution of X under \mathbb{P} . Remind that the measure ν_x constructed in point 1. is the probability distribution of Y given $X = x$ and we will denote by E_x the corresponding mathematical expectation. We will first consider the measure $\nu_x^{(0)}$, the probability distribution of $(Y_t)_{t \geq 0}$ given that $X = x$ and show that the measure $d\gamma(y, x) = d\nu_x^{(0)}(y) d\mu(x)$ is ergodic for the operator $(\theta, \tau).(y, x) = (\theta y, \tau x)$ where for $y \in E^{\mathbb{N}}$, $\theta(y) = (y_{t+1})_{t \in \mathbb{N}}$ and τ has been already defined as the shift operator on $(\mathbb{R}^d)^{\mathbb{Z}}$. For $t \in \mathbb{Z}$, ($t \in \mathbb{N}$ resp.), we denote by y_t the coordinate mapping from $E^{\mathbb{Z}}$ ($E^{\mathbb{N}}$ resp.) to E , i.e. $y_t(z) = z_t$ for $z \in E^{\mathbb{Z}}$ ($z \in E^{\mathbb{N}}$ resp.). Let $B \in \mathcal{P}(E)^{\otimes \mathbb{N}}$, $k \in \mathbb{N}$, n an integer greater than k and $w_0, \dots, w_k \in E$. We have

$$\begin{aligned} \nu_x^{(0)}(y_0 = w_0, \dots, y_k = w_k, y \in \theta^{-n} B) &= \nu_x(y_0 = w_0, \dots, y_k = w_k, (y_t)_{t \geq 0} \in \theta^{-n} B) \\ &= E_x \left[\prod_{i=0}^k \mathbb{1}_{y_i = w_i} \times \nu_x((y_t)_{t \geq 0} \in \theta^{-n} B | y_k^-) \right]. \end{aligned}$$

Using Lemma 1 and the control of the total variation distance mentioned in the point 1. of the Notes, we also have

$$\begin{aligned} |\nu_x((y_t)_{t \geq 0} \in \theta^{-n} B | y_k^-) - \nu_x^{(0)}(\theta^{-n} B)| &\leq \sup_{z, \bar{z}} d_{TV}(\mathcal{Q}_{k,x,z}^{(n-k+1,\infty)}, \mathcal{Q}_{k,x,\bar{z}}^{(n-k+1,\infty)}) \\ &\leq \sum_{i=1}^{\infty} b_{n-k+i-1}^* \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Note also that $\nu_x^{(0)}(\theta^{-n} B) = \nu_{\tau^n x}^{(0)}(B)$. We then get

$$\lim_{n \rightarrow \infty} \sup_{B \in \mathcal{P}(E)^{\otimes \mathbb{N}}} |\nu_x^{(0)}(A \cap \tau^{-n} B) - \nu_x^{(0)}(A) \nu_{\tau^n x}^{(0)}(B)| = 0, \tag{16}$$

when A is a cylinder set. Using approximation by finite unions of disjoint cylinder sets, one can extend (16) to an arbitrary Borel set $A \in \mathcal{P}(E)^{\otimes \mathbb{N}}$. Now let \mathcal{I} be an invariant set in $E^{\mathbb{N}} \times \mathcal{D}$, i.e. $(\theta, \tau)^{-1} \mathcal{I} = \mathcal{I}$. It remains to show that $\gamma(\mathcal{I}) \in \{0, 1\}$. We already mentioned in point 4., the equality $\nu_x^{(0)}(\theta^{-1} A) = \nu_{\tau x}^{(0)}(A)$ when A is a cylinder set. This equality can be extended to any Borel set A . If $\mathcal{I}^x = \{y \in E^{\mathbb{N}} : (y, x) \in \mathcal{I}\}$, we have $\theta^{-1} \mathcal{I}^x = \mathcal{I}^{\tau^{-1} x}$. From (16), we then deduce that

$$\begin{aligned} \nu_x^{(0)}(\mathcal{I}^x) - \nu_x^{(0)}(\mathcal{I}^x)^2 &= \nu_x^{(0)}(\mathcal{I}^x \cap \theta^{-n} \mathcal{I}^{\tau^n x}) - \nu_x^{(0)}(\mathcal{I}^x) \nu_{\tau^n x}^{(0)}(\mathcal{I}^{\tau^n x}) \\ &\rightarrow 0. \end{aligned}$$

This shows that $f(x) := \nu_x^{(0)}(\mathcal{I}^x) \in \{0, 1\}$ for all x . Since $f(\tau x) = f(x)$, ergodicity of X entails that we $\mu(\{f = 1\}) \in \{0, 1\}$ and then $\gamma(\mathcal{I}) \in \{0, 1\}$. This shows that γ is ergodic for (θ, τ) and in particular that the process $(V_t)_{t \in \mathbb{N}}$ is ergodic. But this also entails ergodic

properties for the two-sided sequence $(V_t)_{t \in \mathbb{Z}}$, see for instance Theorem 31 in Douc et al. [16] for a proof.

7.3. Proof of Proposition 1

The result is a consequence of Theorem 1. First we have $b_m \leq L \sum_{j \geq m} |a_j|$ with L the Lipschitz constant of F . Our assumptions entails summability of the b'_m s. The crucial point is to check the condition $b_0 < 1$. Since the first term in the argument of F is bounded, condition $b_0 < 1$ will follow if we show that for any $c > 0$,

$$\sup_{|y| \leq c, z \in \mathbb{R}} |F(y+z) - F(z)| < 1. \tag{17}$$

Note first that F has a limit at $\pm\infty$. Hence, $\sup_{|y| \leq c, |z| > M} F(y+z) < 1/4$ if M is large enough. For such M , we also have $0 < \inf_{|y| \leq c, |z| \leq M} F(y+z) \leq \sup_{|y| \leq c, |z| \leq M} F(y+z) < 1$. We then automatically have (17) and then $b_0 < 1$.

7.4. Proof of Proposition 2

Setting $\lambda_t = (\mu_t, \dots, \mu_{t-q+1})'$, any solution of the problem (9) satisfies the recursions $\lambda_t = A\lambda_{t-1} + b_t$ with

$$A = \begin{pmatrix} \beta_1 & \dots & \beta_q \\ I_{q-1} & & 0_{q-1,1} \end{pmatrix}, \quad b_t = \begin{pmatrix} \sum_{k=1}^p \alpha_k Y_{t-k} + \gamma' X_t \\ 0_{q-1,1} \end{pmatrix},$$

where $0_{q-1,1}$ is a column vector of 0 and I_{q-1} is the identity matrix of size $(q-1) \times (q-1)$. Our assumptions guaranty that the spectral radius of A is less than 1. For a given operator norm $\|\cdot\|$, there then exists $r \in \mathbb{N}^*$ such that $\kappa := \|A^r\| < 1$. One can then apply Lemma 2 to show that any stationary solution $(Y_t)_{t \in \mathbb{Z}}$ satisfying (9) is a chain with complete connections and such that $b_m = O(\kappa^{m/r})$. To finalize the proof, one can apply Theorem 1. One only need to check that $b_0 < 1$. We observe that $q(1|y_{t-1}^-, x_t^-)$ is of the form $F(\sum_{j=1}^\infty \eta_j y_{t-j} + \sum_{k=0}^\infty \delta'_k X_{t-k})$ for some summable sequences $(\eta_j)_{j \geq 1}$ and $(\delta_j)_{j \geq 0}$. Since the first term in the argument of F is bounded, condition $b_0 < 1$ follows exactly as in the proof of Proposition 1, using (17). Theorem 1 entails the result.

7.5. Proof of Proposition 3

As in the proof of Proposition 2, we use Lemma 2 which shows that any solution of (10) is a chain with complete connections for which the coefficients b_m decay geometrically fast. To show S2, it remains to show the condition $b_0 < 1$. Lemma 2 shows that $\sup_{t,x,y,\bar{y}} |\lambda_t^{y,x} - \lambda_t^{\bar{y},x}| = O(1)$. Hence, condition $b_0 < 1$ is implied by (17). Theorem 1 leads to the result.

7.6. Proof of Theorem 2

Using Theorem 1, one can define the process Y conditional on S , instead of conditional on X . The resulting process will be the unique stochastic process satisfies (7). To this end, we simply change the set \mathcal{D} by the set $\mathcal{D}_0 = \{s \in \mathcal{S}^{\mathbb{Z}} : (g(s_t))_{t \in \mathbb{Z}} \in \mathcal{D}\}$. One can then consider that the distribution of the Markov chain S is supported on \mathcal{D}_0 . Secondly, for bounding $\beta_V(n)$ or $\tau_V(n)$, one can replace the sigma-field \mathcal{F}_0 by a larger one. This follows from the properties of the conditional expectations. Let μ be the probability distribution of S_0^- . We also denote by $K_{s_0^-}$ the probability distribution of Y_0^- conditional on $S = s$. Note that from Theorem 1, this conditional distribution only depends on s_0^- .

For $(t, w, z, s, \bar{s}) \in \mathbb{Z} \times E \times E^{\mathbb{N}} \times \mathcal{D}_0 \times \mathcal{D}_0$ and $x = x(s, \bar{s}) = ((g(s_t))_{t \in \mathbb{Z}}, (g(\bar{s}_t))_{t \in \mathbb{Z}})$, we set

$$q_t^x(w|z) = q(w|z, g(s_t)^-), \quad \bar{q}_t^x(w|z) = q(w|z, g(\bar{s}_t)^-).$$

Next setting $s_1^+ = (s_1, s_2, \dots)$ and $y_1^+ = (y_1, y_2, \dots)$ for any $(s, y) \in \mathcal{D}_0 \times E^{\mathbb{N}}$, we consider a probability measure \mathbb{P} on $\Omega = \mathcal{S}^{\mathbb{Z}} \times \mathcal{S}^{\mathbb{Z}} \times E^{\mathbb{Z}} \times E^{\mathbb{Z}}$ endowed with its Borel σ -field and defined by

$$\begin{aligned} &\mathbb{P}(ds, d\bar{s}, dy, d\bar{y}) \\ &= \mu(ds_0^-)\mu(d\bar{s}_0^-)K_{s_0^-}(dy_0^-)K_{\bar{s}_0^-}(d\bar{y}_0^-)\tilde{P}((s_0, \bar{s}_0), (ds_1^+, d\bar{s}_1^+))\tilde{Q}_{0,x(s,\bar{s}),z,\bar{z}}(dy_1^+, d\bar{y}_1^+). \end{aligned}$$

We remind that \tilde{P} is defined in Assumption S1' and \tilde{Q} is defined in Lemma 1. On Ω , we will still denote, for $t \in \mathbb{Z}$, the coordinate applications by $Y_t, \bar{Y}_t, S_t, \bar{S}_t$. Let us also point out that the measure

$$\tilde{P}((s_0, \bar{s}_0), (ds_1^+, d\bar{s}_1^+))\tilde{Q}_{0,x(s,\bar{s}),z,\bar{z}}(dy_1^+, d\bar{y}_1^+)$$

is a coupling of two conditional distributions, the distribution of $((Y_t, S_t))_{t \geq 1}$ conditional on $Y_j = y_j, S_j = s_j$ for $j \leq 0$ and the distribution of $((\bar{Y}_t, \bar{S}_t))_{t \geq 1}$ conditional on $Y_j = \bar{y}_j, S_j = \bar{s}_j$ for $j \leq 0$. Next we set $\mathcal{G}_0 = \sigma((Y_j, S_j) : j \leq 0)$ and $\bar{\mathcal{G}}_0 = \sigma((\bar{Y}_j, \bar{S}_j) : j \leq 0)$. Let $J = \{j_1, \dots, j_\ell\} \subset \mathbb{N}^*$. Note that the two sigma-fields \mathcal{G}_0 and $\bar{\mathcal{G}}_0$ are independent.

1. For the absolute regularity coefficients, we use the bounds

$$\begin{aligned} \beta_V(n) &\leq \mathbb{E}\left[\sup_A |\mathbb{P}((V_n, V_{n+1}, \dots) \in A | \mathcal{G}_0) - \mathbb{P}((\bar{V}_n, \bar{V}_{n+1}, \dots) \in A | \bar{\mathcal{G}}_0)|\right] \\ &\leq \sum_{t \geq n} \mathbb{E}[\mathbb{P}(V_n \neq \bar{V}_n | \mathcal{G}_0 \vee \bar{\mathcal{G}}_0)]. \end{aligned}$$

Next, using Lemma 1 and Assumption S3, we have

$$\begin{aligned} \mathbb{P}(Y_t \neq \bar{Y}_t | \mathcal{G}_0 \vee \bar{\mathcal{G}}_0 \vee \sigma(S_1^+, \bar{S}_1^+)) &\leq \sup_{z_0^-, \bar{z}_0^-} \tilde{Q}_{0,x(s,\bar{s}),z_0^-, \bar{z}_0^-}(\{y_t \neq \bar{y}_t\}) \\ &\leq b_{t-1}^* + \sup_{g \in E^{\mathbb{N}}} d_{TV}(q(\cdot | g, S_t^-), q(\cdot | g, \bar{S}_t^-)) \end{aligned}$$

$$\begin{aligned}
& + \sum_{\ell=0}^{t-2} b_{\ell}^* \sup_{g \in E^{\mathbb{N}}} d_{\text{TV}}(q(\cdot|g, S_{t-\ell-1}^{-}), q(\cdot|g, \bar{S}_{t-\ell-1}^{-})) \\
& \leq b_{t-1}^* + \sum_{i \geq 0} e_i G_{t-i} + \sum_{\ell=0}^{t-2} b_{\ell}^* \sum_{i=0}^{\infty} e_i G_{t-\ell-i-1},
\end{aligned}$$

where for any $t \in \mathbb{Z}$, $G_t = |g(S_t) - g(\bar{S}_t)|$. Using Holder inequality, we have

$$\mathbb{E}(G_t) = \mathbb{E}[|g(S_t) - g(\bar{S}_t)| \mathbb{1}_{g(S_t) \neq g(\bar{S}_t)}] \leq 2 \|X_0\|_p \mathbb{P}(g(S_t) \neq g(\bar{S}_t))^{1/q}.$$

From the definition of the coupling, we have $\mathbb{P}(g(S_t) \neq g(\bar{S}_t)) = a_t$ where a_t is defined in S1'. Since,

$$\mathbb{P}(V_n \neq \bar{V}_n | \mathcal{G}_0 \vee \bar{\mathcal{G}}_0) \leq \mathbb{P}(Y_t \neq \bar{Y}_t | \mathcal{G}_0 \vee \bar{\mathcal{G}}_0) + \mathbb{P}(g(S_t) \neq g(\bar{S}_t) | \sigma(S_0, \bar{S}_0))$$

and $a_t = \mathbb{P}(g(S_t) \neq g(\bar{S}_t)) \leq c_t$, the bound for $\beta_V(n)$ follows after integrating the previous inequalities.

2. We have

$$\tau(\mathcal{F}_0, U_J) \leq \mathbb{E}[W_{1,\ell}(\mathbb{P}_{U_J|\mathcal{F}_0}, \mathbb{P}_{\bar{U}_J|\bar{\mathcal{F}}_0})],$$

with $\bar{U}_J = (\bar{V}_{j_1}, \dots, \bar{V}_{j_\ell})$ and $\bar{V}_t = (\bar{Y}_t, g(\bar{S}_t))$ for $t \in \mathbb{Z}$. Using our coupling we have

$$\begin{aligned}
& W_{1,\ell}(\mathbb{P}_{U_J|\mathcal{G}_0}, \mathbb{P}_{\bar{U}_J|\bar{\mathcal{G}}_0}) \\
& \leq \sum_{i=1}^{\ell} [\mathbb{P}(Y_{j_i} \neq \bar{Y}_{j_i} | \mathcal{G}_0 \vee \bar{\mathcal{G}}_0) + \mathbb{E}(|g(S_{j_i}) - g(\bar{S}_{j_i})| | \sigma(S_0, \bar{S}_0))] \\
& \leq \ell \sup_{t \geq n} [\mathbb{P}(Y_t \neq \bar{Y}_t | \mathcal{G}_0 \vee \bar{\mathcal{G}}_0) + \mathbb{E}(|g(S_t) - g(\bar{S}_t)| | \sigma(S_0, \bar{S}_0))].
\end{aligned}$$

From the definition of our coupling and Assumption S1', we have

$$\mathbb{E}[|g(S_t) - g(\bar{S}_t)|] \leq a_t.$$

Next, one can bound $\mathbb{P}(Y_t \neq \bar{Y}_t)$ as in the previous point and we have directly $\mathbb{E}(G_t) = a_t$. The proposed upper-bound for $\tau_V(n)$ easily follows.

7.7. Proof of Corollary 2

From Lemma 2, the coefficients b_i decay exponentially and from Lemma 4 given in the supplementary material, so do the corresponding coefficients b_i^* . Moreover, the coefficients e_i in Assumption S3 also decay exponentially fast. For a polynomial decay, we have $g_j = O(j^{-\kappa/q})$ and $h_j = O(j^{-\kappa})$ in Theorem 2. The result of the corollary then follows.

Appendix

The proof of the following lemma can be found in the supplementary material.

Lemma 2. *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary process taking values in \mathbb{R}^d , $(y_t)_{t \in \mathbb{Z}}$ a sequence of point in E and $\{G_{y,x} : (y, x) \in E \times \mathbb{R}^d\}$ a family of applications from \mathbb{R}^k to \mathbb{R}^k satisfying the three following assumptions.*

1. *There exists $L \geq 1$ such that for $(y, y', x, x', z, z') \in E^2 \times \mathbb{R}^{2d} \times \mathbb{R}^{2k}$,*

$$|G_{y,x}(z) - G_{y',x'}(z')| \leq L[\mathbb{1}_{y \neq y'} + |x - x'| + |z - z'|].$$

2. *There exist a positive integer r and a real number $\kappa \in (0, 1)$ such that for all $(y_1, \dots, y_r) \in E^r$, $(x_1, \dots, x_r) \in \mathbb{R}^{dr}$ and $(s, s') \in \mathbb{R}^{2k}$,*

$$|G_{y_1, x_1} \circ \dots \circ G_{y_r, x_r}(s) - G_{y_1, x_1} \circ \dots \circ G_{y_r, x_r}(s')| \leq \kappa |s - s'|.$$

3. $\mathbb{E} \log^+ |X_1| < \infty$.

Then the following conclusions hold true.

- *Setting $\mathcal{D} = \{x \in (\mathbb{R}^d)^\mathbb{N} : \sum_{i=0}^\infty \kappa^{i/r} |x_i| < \infty\}$, we have $\mathbb{P}(X \in \mathcal{D}) = 1$.*
- *For $t \in \mathbb{Z}$, $n \in \mathbb{N}^*$, $y \in E^\mathbb{Z}$ and $x \in \mathcal{D}$, set $\lambda_{n,t}^{(y,x)} = G_{y_{t-1}, x_t} \circ \dots \circ G_{y_{t-n-1}, x_{t-n}}(0)$. Then for any $t \in \mathbb{Z}$, the sequence $(\lambda_{n,t}^{(y,x)})_{n \geq 1}$ converges to an element of \mathbb{R}^k denoted by $\lambda_t^{(y,x)}$. Moreover, there exists $H : E^\mathbb{N} \times \mathcal{D} \rightarrow \mathbb{R}^k$ such that $\lambda_t = H(y_{t-1}^-, x_t^-)$.*
- *Let $y \in E^\mathbb{Z}$ and $x \in (\mathbb{R}^k)^\mathbb{Z}$. If $(\bar{\lambda}_t)_{t \in \mathbb{Z}}$ is a sequence in \mathbb{R}^k , such that $\liminf_{t \rightarrow -\infty} |\bar{\lambda}_t| < \infty$ and $\bar{\lambda}_t = G_{y_{t-1}, x_t}(\bar{\lambda}_{t-1})$ for all $t \in \mathbb{Z}$, then $\bar{\lambda}_t = \lambda_t^{(y,x)}$ for all $t \in \mathbb{Z}$, where $\lambda_t^{(x,y)}$ is defined in the previous question.*
- *Keeping the notations given in the previous point, we have for an integer $m \geq 1$*

$$\sup_{x \in \mathcal{D}} \sup_{y_i = y'_i, -m+1 \leq i \leq 0} |\lambda_0^{(y,x)} - \lambda_0^{(y',x)}| = O(\kappa^{m/r}).$$

- *We have*

$$|H(y_{t-1}^-, x_t^-) - H(y_{t-1}^-, \bar{x}_t^-)| \leq \sum_{j=0}^\infty \kappa^j \sum_{i=1}^r L^i |x_{t-jr-i+1} - \bar{x}_{t-jr-i+1}|.$$

- *We assume that $((Y_t, X_t))_{t \in \mathbb{Z}}$ is a stationary process taking values in $E \times \mathbb{R}^d$. Then a stationary process $(\lambda_t)_{t \in \mathbb{Z}}$ satisfies the recursions $\lambda_t = G_{Y_{t-1}, X_t}(\lambda_{t-1})$ if and only if $\lambda_t = H(Y_{t-1}^-, X_t^-)$, where H is defined in the second point.*

Supplementary Material

Supplement to “Coupling and perturbation techniques for categorical time series” (DOI: 10.3150/20-BEJ1225SUPP; .pdf). We provide additional proofs that are not given in the paper.

References

- [1] Bressaud, X., Fernández, R. and Galves, A. (1999). Decay of correlations for non-Hölderian dynamics. A coupling approach. *Electron. J. Probab.* **4** 3. MR1675304 <https://doi.org/10.1214/EJP.v4-40>
- [2] Bressaud, X., Fernández, R. and Galves, A. (1999). Speed of \bar{d} -convergence for Markov approximations of chains with complete connections. A coupling approach. *Stochastic Process. Appl.* **83** 127–138. MR1705603 [https://doi.org/10.1016/S0304-4149\(99\)00025-3](https://doi.org/10.1016/S0304-4149(99)00025-3)
- [3] Busch, J.R., Ferrari, P.A., Flesia, A.G., Fraiman, R., Grynberg, S.P. and Leonardi, F. (2009). Testing statistical hypothesis on random trees and applications to the protein classification problem. *Ann. Appl. Stat.* **3** 542–563. MR2750672 <https://doi.org/10.1214/08-AOAS218>
- [4] Candelon, B., Dumitrescu, E.-I., Hurlin, C. and Palm, F.C. (2013). Multivariate dynamic probit models: An application to financial crises mutation. In *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims. Adv. Econom.* **32** 395–427. Bingley: Emerald Group Publ. Ltd. MR3496842 <https://doi.org/10.1108/S0731-905320130000031011>
- [5] Chamberlain, G. (1982). The general equivalence of Granger and Sims causality. *Econometrica* **50** 569–581. MR0662719 <https://doi.org/10.2307/1912601>
- [6] Chazottes, J.-R., Gallo, S. and Takahashi, D. (2020). Optimal gaussian concentration bounds for stochastic chains of unbounded memory. ArXiv preprint. Available at [arXiv:2001.06633](https://arxiv.org/abs/2001.06633).
- [7] Cogburn, R. (1984). The ergodic theory of Markov chains in random environments. *Z. Wahrsch. Verw. Gebiete* **66** 109–128. MR0743088 <https://doi.org/10.1007/BF00532799>
- [8] Comets, F., Fernández, R. and Ferrari, P.A. (2002). Processes with long memory: Regenerative construction and perfect simulation. *Ann. Appl. Probab.* **12** 921–943. MR1925446 <https://doi.org/10.1214/aoap/1031863175>
- [9] Cox, D.R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K. and Lauritzen, S.L. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scand. J. Stat.* 93–115.
- [10] de Jong, R.M. and Woutersen, T. (2011). Dynamic time series binary choice. *Econometric Theory* **27** 673–702. MR2822362 <https://doi.org/10.1017/S0266466610000472>
- [11] Dedecker, J., Doukhan, P., Lang, G., León R., J.R., Louhichi, S. and Prieur, C. (2007). *Weak Dependence: With Examples and Applications. Lecture Notes in Statistics* **190**. New York: Springer. MR2338725
- [12] Dedecker, J. and Prieur, C. (2004). Coupling for τ -dependent sequences and applications. *J. Theoret. Probab.* **17** 861–885. MR2105738 <https://doi.org/10.1007/s10959-004-0578-x>
- [13] Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields* **132** 203–236. MR2199291 <https://doi.org/10.1007/s00440-004-0394-3>
- [14] den Hollander, F. (2012). *Probability Theory: The Coupling Method. Lecture Notes*. Available at <http://websites.math.leidenuniv.nl/probability/lecturenotes/CouplingLectures.pdf>.
- [15] Doeblin, W. and Fortet, R. (1937). Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France* **65** 132–148. MR1505076
- [16] Douc, R., Doukhan, P. and Moulines, E. (2013). Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stochastic Process. Appl.* **123** 2620–2647. MR3054539 <https://doi.org/10.1016/j.spa.2013.04.010>
- [17] Doukhan, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statistics* **85**. New York: Springer. MR1312160 <https://doi.org/10.1007/978-1-4612-2642-0>
- [18] Fernández, R. and Galves, A. (2002). Markov approximations of chains of infinite order. *Bull. Braz. Math. Soc. (N.S.)* **33** 295–306. MR1978829 <https://doi.org/10.1007/s005740200015>

- [19] Fernández, R. and Maillard, G. (2005). Chains with complete connections: General theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.* **118** 555–588. MR2123648 <https://doi.org/10.1007/s10955-004-8821-5>
- [20] Fokianos, K. and Kedem, B. (2003). Regression theory for categorical time series. *Statist. Sci.* **18** 357–376. MR2061915 <https://doi.org/10.1214/ss/1076102425>
- [21] Fokianos, K., Rahbek, A. and Tjøstheim, D. (2009). Poisson autoregression. *J. Amer. Statist. Assoc.* **104** 1430–1439. MR2596998 <https://doi.org/10.1198/jasa.2009.tm08270>
- [22] Fokianos, K. and Truquet, L. (2019). On categorical time series models with covariates. *Stochastic Process. Appl.* **129** 3446–3462. MR3985569 <https://doi.org/10.1016/j.spa.2018.09.012>
- [23] Galves, A., Galves, C., García, J.E., Garcia, N.L. and Leonardi, F. (2012). Context tree selection and linguistic rhythm retrieval from written texts. *Ann. Appl. Stat.* **6** 186–209. MR2951534 <https://doi.org/10.1214/11-AOAS511>
- [24] Guanche, Y., Mínguez, R. and Méndez, F.J. (2014). Autoregressive logistic regression applied to atmospheric circulation patterns. *Clim. Dyn.* **42** 537–552.
- [25] Hao, Z., Hao, F., Xia, Y., Singh, V.P., Hong, Y., Shen, X. and Ouyang, W. (2016). A statistical method for categorical drought prediction based on nlds-2. *J. Appl. Meteorol. Climatol.* **55** 1049–1061.
- [26] Harris, T.E. (1955). On chains of infinite order. *Pacific J. Math.* **5** 707–724. MR0075482
- [27] Iosifescu, M. and Grigorescu, Ş. (1990). *Dependence with Complete Connections and Its Applications. Cambridge Tracts in Mathematics* **96**. Cambridge: Cambridge Univ. Press. MR1070097
- [28] Kaufmann, H. (1987). Regression models for nonstationary categorical time series: Asymptotic estimation theory. *Ann. Statist.* **15** 79–98. MR0885725 <https://doi.org/10.1214/aos/1176350254>
- [29] Kauppi, H. and Saikkonen, P. (2008). Predicting us recessions with dynamic binary response models. *Rev. Econ. Stat.* **90** 777–791.
- [30] Kifer, Y. (1996). Perron–Frobenius theorem, large deviations, and random perturbations in random environments. *Math. Z.* **222** 677–698. MR1406273 <https://doi.org/10.1007/PL00004551>
- [31] Kifer, Y. (1998). Limit theorems for random transformations and processes in random environments. *Trans. Amer. Math. Soc.* **350** 1481–1518. MR1451607 <https://doi.org/10.1090/S0002-9947-98-02068-6>
- [32] Lindvall, T. (2002). *Lectures on the Coupling Method*. New York: Dover.
- [33] Manner, H., Türk, D. and Eichler, M. (2016). Modeling and forecasting multivariate electricity price spikes. *Energy Econ.* **60** 255–265.
- [34] Merlevède, F., Peligrad, M. and Rio, E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Related Fields* **151** 435–474. MR2851689 <https://doi.org/10.1007/s00440-010-0304-9>
- [35] Mitrophanov, A.Y. (2005). Sensitivity and convergence of uniformly ergodic Markov chains. *J. Appl. Probab.* **42** 1003–1014. MR2203818 <https://doi.org/10.1239/jap/1134587812>
- [36] Neumann, M.H. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **17** 1268–1284. MR2854772 <https://doi.org/10.3150/10-BEJ313>
- [37] Nyberg, H. (2014). A bivariate autoregressive probit model: Business cycle linkages and transmission of recession probabilities. *Macroecon. Dyn.* **18** 838–862.
- [38] Park, B.U., Simar, L. and Zelenyuk, V. (2017). Nonparametric estimation of dynamic discrete choice models for time series data. *Comput. Statist. Data Anal.* **108** 97–120. MR3589904 <https://doi.org/10.1016/j.csda.2016.10.024>
- [39] Rao, P. (2009). Conditional independence, conditional mixing and conditional association. *Ann. Inst. Statist. Math.* **61** 441–460. MR2505397 <https://doi.org/10.1007/s10463-007-0152-2>
- [40] Rissanen, J. (1983). A universal data compression system. *IEEE Trans. Inf. Theory* **29** 656–664. MR0730903 <https://doi.org/10.1109/TIT.1983.1056741>

- [41] Russell, J.R. and Engle, R.F. (2005). A discrete-state continuous-time model of financial transactions prices and times: The autoregressive conditional multinomial – autoregressive conditional duration model. *J. Bus. Econom. Statist.* **23** 166–180. MR2157268 <https://doi.org/10.1198/073500104000000541>
- [42] Rydberg, T.H. and Shephard, N. (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *J. Financ. Econom.* **1** 2–25.
- [43] Sims, C.A. (1972). Money, income, and causality. *Am. Econ. Rev.* **62** 540–552.
- [44] Truquet, L. (2020). Supplement to “Coupling and perturbation techniques for categorical time series.” <https://doi.org/10.3150/20-BEJ1225SUPP>
- [45] Villani, C. (2009). *Optimal Transport. Old and New*. Berlin: Springer.
- [46] Woodard, D.B., Matteson, D.S. and Henderson, S.G. (2011). Stationarity of generalized autoregressive moving average models. *Electron. J. Stat.* **5** 800–828. MR2824817 <https://doi.org/10.1214/11-EJS627>
- [47] Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 <https://doi.org/10.1073/pnas.0506715102>
- [48] Yuan, D. and Lei, L. (2013). Some conditional results for conditionally strong mixing sequences of random variables. *Sci. China Math.* **56** 845–859. MR3034846 <https://doi.org/10.1007/s11425-012-4554-0>

Received September 2019 and revised April 2020