LOCAL STATIONARITY AND TIME-INHOMOGENEOUS MARKOV CHAINS

By Lionel Truquet ,

University Rennes 1 *and ENSAI*

A primary motivation of this contribution is to define new locally stationary Markov models for categorical or integer-valued data. For this initial purpose we propose a new general approach for dealing with time-inhomogeneity that extends the local stationarity notion developed in the time series literature. We also introduce a probabilistic framework which is very flexible and allows us to consider a much larger class of Markov chain models on arbitrary state spaces, including most of the locally stationary autoregressive processes studied in the literature. We consider triangular arrays of time-inhomogeneous Markov chains, defined by some families of contracting and slowly-varying Markov kernels. The finite-dimensional distribution of such Markov chains can be approximated locally with the distribution of ergodic Markov chains and some mixing properties are also available for these triangular arrays. As a consequence of our results, some classical geometrically ergodic homogeneous Markov chain models have a locally stationary version, which lays the theoretical foundations for new statistical modeling. Statistical inference of finite-state Markov chains can be based on kernel smoothing and we provide a complete and fast implementation of such models, directly usable by the practitioners. We also illustrate the theory on a real data set. A central limit theorem for Markov chains on more general state spaces is also provided and illustrated with the statistical inference in INAR models, Poisson ARCH models and binary time series models. Additional examples such as locally stationary regime-switching or SETAR models are also discussed.

1. Introduction. Markov chains are one of the most basic examples of random sequences used for the statistical modeling of dependent data. For instance, finite-state Markov chains have important applications in queuing systems (see for instance Bolch et al. [7]), for the modeling of DNA sequences (see Avery and Henderson [3]) or in computer networks (see for instance Sarukkai [42]). Moreover, some classical time series models based on AR or ARCH processes can be seen as particular examples of Markov chains on a continuous state space. Other examples concern integer-valued time series modes such as the INAR process introduced by Al Osh and Alzaid

1

[1] or the Poisson autoregressive process studied for instance in Fokianos et al. [24]. Another type of Markov chain widely encountered in time series analysis is the Markov-switching process introduced initially by Hamilton [30] in economics.

However, a crucial limitation of Markov chain models is their time homogeneity. In practice, this assumption is violated for many data sets. For instance, analyzing the trading activity of a traded share on the Johannesburg Stock Exchange (the data are binary and the sequence takes the value 1 if a trading has been recorded at time t), we have estimated locally the probability of recording a trade at time t for two traded shares (see Sections 5 and 6.2 for details on the data set and for the smoothing used for this estimation). Figure 1 suggests that the marginal distributions of both sequences are far from being time-invariant.
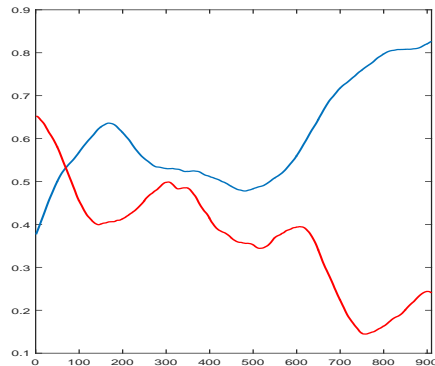

FIG 1. *Local estimation of the probability to record a trade for the share "Anamint" (in blue) and "Broadcares" (in red). The $x-$axis represents the time and the $y-$axis, the estimated local probability.*

Motivated by this type of problem, in this paper, we develop a probabilistic framework for defining time-inhomogeneous Markov chain models on arbitrary state spaces and which is can used for statistical inference. Time-inhomogeneous Markov chains have received much less attention in the literature than their homogeneous analogues. Such chains have been studied mainly for their long-time behavior, often in connection with the convergence of stochastic algorithms. An introduction to inhomogeneous finite-state Markov chains can be found in Seneta [43] and their use in Monte Carlo methods is discussed in Winkler [51]. More recent quantitative results for their long-time behavior can be found for instance in Douc et al. [19] for general state spaces, and Saloff-Coste and Zúñiga [40] or Saloff-Coste

and Zúñiga [41] for finite state spaces. In this paper, we consider inhomogeneous Markov chain models for applications in statistics, in the spirit of the notion of local stationarity introduced by Dahlhaus [12]. Locally stationary processes have received considerable attention over the last twenty years, in particular for their ability to model data sets for which time-homogeneity is unrealistic. Locally stationary, autoregressive processes (here with one lag for simplicity) can be defined by modifying a recursive equation followed by a stationary process. If $(X_k)_{k\in\mathbb{Z}}$ is a stationary process defined by $X_k = F_\theta(X_{k-1}, \varepsilon_k)$, where $(\varepsilon_k)_{k\in\mathbb{Z}}$ is a sequence of i.i.d. random variables and $\theta \in \Theta$ is a parameter, its locally stationary version is usually defined recursively by

$$(1) \qquad X_{n,k} = F_{\theta(k/n)}(X_{n,k-1}, \varepsilon_k), \quad 1 \le k \le n,$$

where $\theta : [0,1] \to \Theta$ is a smooth function. This formalism was exploited for defining locally stationary versions of classical time-homogeneous, autoregressive processes. See for instance Dahlhaus and Subba Rao [13], Subba Rao [45] or Vogt [50]. The term local stationarity comes from the fact that, under some regularity conditions, if k/n is close to a point u of $[0,1]$, $X_{n,k}$ is close in some sense to $X_k(u)$ where $(X_k(u))_{k\in\mathbb{Z}}$ is the stationary process defined by

$$(2) \qquad X_k(u) = F_{\theta(u)}(X_{k-1}(u), \varepsilon_k), \quad k \in \mathbb{Z}.$$

General Markov models of this type are considered in the recent paper of Dahlhaus et al. [14]. However, the models we will consider in this paper can have a quite different structure. They can be defined by a conditional distribution (finite-state Markov chains, Poisson ARCH processes, Logistic autoregressive processes), the autoregressive representation may have different shape (INAR processes) or some discontinuities (SETAR models). It is then unclear how to define local stationarity in this context. Note also that the current approach for defining locally stationary Markov chains consists in defining a family of autoregressive processes adapted the same sequence of innovations $(\varepsilon_t)_{t\in\mathbb{Z}}$. A general result, see for instance Theorem 5.24 in Douc et al. [20], asserts that on general state spaces, every Markov chain can be represented by an autoregressive process. However, the function F will not display any useful property in general. Moreover, existing constructions of locally stationary models such as in Dahlhaus et al. [14] or Vogt [50] assume some smoothness properties for the random function $(u, x) \mapsto F_{\theta(u)}(x, \varepsilon_1)$. Clearly, this approach is not adapted to finite-state Markov chains and then to more general models such as Markov switching. Moreover, even with a

natural autoregressive representation, the contraction property in $\mathbb{L}^p-$norms required in Dahlhaus et al. [14] is invalid for stationary SETAR processes.

On the other hand, some properties of time-homogeneous Markov chains are derived more easily by analytical methods. For instance the ergodicity properties of finite-state Markov chains are easily derived form the transition matrix and this example was a motivation to define a local stationarity notion by using properties of the probability distributions instead an almost sure autoregressive representation.

Let us now give the framework used in the rest of the paper. Let (E, d) be a metric space, $\mathcal{B}(E)$ its corresponding Borel $\sigma-$field and $\{Q_u : u \in [0, 1]\}$ a family of Markov kernels on $(E, \mathcal{B}(E))$. By convention, we will assume that $X_{n,0}$ has the probability distribution π_0. We remind the reader that a Markov kernel $R : E \times \mathcal{B}(E) \to [0, 1]$ on $(E, \mathcal{B}(E))$ is an application such that for all $(y, A) \in E \times \mathcal{B}(E)$, the application $x \mapsto R(x, A)$ is measurable and the application $A \mapsto R(x, A)$ defines a probability measure. We will consider triangular arrays $\{X_{n,j} : 1 \leq j \leq n, n \in \mathbb{Z}^+\}$ such that for all $n \in \mathbb{Z}^+$, the sequence $(X_{n,j})_{1 \leq j \leq n}$ is a non homogeneous Markov chain such that

$$\mathbb{P}(X_{n,k} \in A | X_{n,k-1} = x) = Q_{k/n}(x, A), \quad 1 \leq k \leq n.$$

The family $\{Q_u : u \in [0, 1]\}$ of Markov kernels will always satisfy some regularity conditions and contraction properties. Precise assumptions will be given in the three following sections, but from now on, we assume here that for all $u \in [0, 1]$, Q_u has a single invariant probability measure denoted by π_u. It could be also convenient to define all the random variables $X_{n,k}$ on the same probability space, in particular for different values of the integer n. However, the dependence of two Markov chains with distinct values of n has no importance in our study and one can simply assume mutual independence between the lines of this triangular array.

For all positive integers j and k such that $k + j - 1 \leq n$, we denote by $\pi_{k,j}^{(n)}$ the probability distribution of the vector $(X_{n,k}, X_{n,k+1}, \ldots, X_{n,k+j-1})$ and by $\pi_{u,j}$ the corresponding finite dimensional distribution for the ergodic chain with Markov kernels Q_u. Loosely speaking, the triangular array will be said to be locally stationary if for all positive integer j, the probability distribution $\pi_{k,j}^{(n)}$ is close to $\pi_{u,j}$ when the ratio k/n is close to u. For compatibility and simplicity of our notations, the measures $\pi_{k,1}^{(n)}$ (resp. $\pi_{u,1}$) will be simply denoted by $\pi_k^{(n)}$ (resp. π_u). A formal definition is given below. For all integers $j \geq 1$, we denote by $\mathcal{P}(E^j)$ the set of probability measures on $(E^j, \mathcal{B}(E^j))$.

DEFINITION 1.   *Let $\vartheta = (\vartheta_j)_{j \geq 1} \in \prod_{j \geq 1} \mathcal{P}(E^j)$ be a sequence of probability measures which define a stronger topology than weak convergence. The triangular array of inhomogeneous Markov chains $\{X_{n,k}, n \in \mathbb{Z}^+, 1 \leq k \leq n\}$ is said to be $\vartheta-$ locally stationary if the two following conditions are satisfied.*

1. *For all $j \geq 1$, the application $u \mapsto \pi_{u,j}$ is continuous from $([0,1], |\cdot|)$ to $\left(\mathcal{P}(E^j), \vartheta_j\right)$.*
2. *For all $j \geq 1$, $\lim_{n \to \infty} \sup_{1 \leq k \leq n-j+1} \vartheta_j(\pi_{k,j}^{(n)}, \pi_{\frac{k}{n}, j}) = 0$.*

Note that Definition 1 is not restricted to Markov chains and can be used as a general definition for local stationarity of time series. In this case, the $\pi_{u,j}$'s are simply the finite dimensional distributions (f.d.d. in the sequel) of some stationary time series. Under the two conditions of Definition 1, for all continuous and bounded functions $f : E^j \to \mathbb{R}$ and some integers $1 \leq k = k_n \leq n - j + 1$ such that $\lim_{n \to \infty} k/n = u \in [0,1]$, we have

$$\lim_{n \to \infty} \mathbb{E}f\left(X_{n,k}, \ldots, X_{n,k+j-1}\right) = \mathbb{E}f\left(X_1(u), \ldots, X_j(u)\right) = \int f d\pi_{u,j},$$

where $(X_k(u))_{k \in \mathbb{Z}}$ denotes a stationary Markov chain with transition kernel Q_u. Note that the coordinates of this last Markov chain are defined for convenience but the process need not to be defined on the same probability space as the triangular array. Definition 1 gives minimal conditions for defining triangular arrays of random variables for which the f.d.d. are locally approximable by stationary processes with continuously changing f.d.d. However, this definition is not sufficient for statistical inference. Making a parallel with stationary processes, the simple definition of stationarity is not sufficient for constructing a valid asymptotic theory and mixing type conditions for the stochastic process are needed. Additionally, for locally stationary processes, a rate is needed in point 2. of Definition 1 as well as some regularity assumptions for the functions $u \mapsto \int f d\pi_{u,j}$, $j \geq 1$, for example Lipschitz continuity or existence of some derivatives.

Another important issue is to find some suitable probabilistic metrics ϑ_j for which Definition 1 is satisfied. Of course, the metrics ϑ_j will be of the same nature for different integers j, e.g. the total variation distance on $\mathcal{P}(E^j)$. In view of the nonparametric estimation of the transition kernel $u \mapsto Q_u$, it is necessary to obtain a convergence rate in the condition 2 of Definition 1. In this paper, this rate will be obtained using a contraction property for the Markov kernels function Q_u (sometimes after iteration) as well as Lipschitz continuity of the application $u \mapsto Q_u$ for the metric ϑ_1. Interestingly, the contraction properties will also guarantee some mixing

properties for the triangular array and its stationary approximations at the same time. A direct consequence of these properties is a control of the bias and the variance of localized partial sums of the process, which is basis for deriving asymptotic properties of localized minimum of contrast estimators. See for instance Theorem 4 in Section 4.

In the literature, there exists another recent contribution that defines local stationarity using probability distances. In Birr et al. [6], the authors use the Kolmogorov-Smirnov distance to control the approximation of a locally stationary process by a stationary one. However this metric is only interesting for finite-dimensional state-spaces and more natural for $E = \mathbb{R}^d$. Since this metric will not provide additional examples of locally stationary Markov chains, we will not use it and we will focus in this paper on three types of metrics of classical use for studying geometric ergodicity properties of Markov chains. To this end, we will extensively make use of the so-called Dobrushin's contraction coefficient.

Let us also mention that our approach provides a rigorous framework to some previous contributions devoted to the fitting of time-inhomogeneous, finite-state Markov chains. The approach of Vergne [48] for modeling DNA sequences, Rajagopalan et al. [38] in hydrology or Brillinger et al. [10], are also closely related to the concept of local stationarity but no precise statistical model is introduced to support the applications considered in these papers. In a different context, one can also mention the work of Hall and Bura [29] which is devoted to the nonparametric estimation of nonhomogeneous continuous time Markov process sampled at i.i.d. random times.

The paper is organized as follows. In Section 2, we consider the total variation distance. This is the metric for which the contraction coefficient for Markov kernels has been introduced originally by Dobrushin [17]. Contraction properties of the kernels Q_u or their iteration with respect to this metric is mainly adapted to compact state spaces and will enable us to consider a model of inhomogeneous finite-state space Markov chains for which we will study a nonparametric estimator of the time-varying transition matrix. Our results in total variation apply when all the Markov kernels Q_u are absolutely continuous with respect to a given reference measure.

In Section 3, we consider Markov kernels contracting in Wasserstein metrics. The contraction coefficient for the Wasserstein metric of order 1 has been first considered by Dobrushin [18] for giving sufficient conditions under which a system of conditional distributions defines a unique joint distribution. Our goal in this section will be simply to show that many autoregressive processes with time-varying coefficients considered in the literature satisfy our assumptions and can be seen as particular examples of our general ap-

proach. In Section 4, we extend the results of Section 1 by considering some Markov kernels which satisfy classical drift/small set conditions. The results of Section 2 can be deduced from the results of Section 4. However for the reader's convenience, we present a separate result for compact state spaces with sharper mixing properties. Our third approach is illustrated with several new examples of locally stationary processes, including INAR processes, Markov switching autoregressive processes and SETAR models. We also discuss statistical inference for some of these models, such as INAR, Poisson ARCH and binary time series, using local least-squares or local likelihood estimators. Section 5 is devoted to the practical implementation of finite-state Markov chains and in Section 6, we consider an illustration on a real data set. A discussion of our results and a guideline for applying them is given in Section 7. The proofs of all our results are available in the supplementary material which also contains additional examples of locally stationary Markov chains as well as a discussion of the mixing properties of the Markov chains studied in Section 3.

2. Total variation distance and finite-state space Markov chains. This section is mainly motivated by finite-state Markov chains. The limiting behavior of finite-state Markov chains is often studied using the total variation distance. See for instance Seneta [43] or Winkler [51]. However, the result stated in this section can be used to construct locally stationary Markov chains on more general compact state spaces. Let us first introduce some notations that we will extensively use in the rest of the paper. If $\mu \in \mathcal{P}(E)$ and R is a probability kernel from $(E, \mathcal{B}(E))$ to $(E, \mathcal{B}(E))$, we will denote by μR the probability measure defined by

$$\mu R(A) = \int R(x, A) d\mu(x), \quad A \in \mathcal{B}(E).$$

Moreover if $f : E \to \mathbb{R}$ is a measurable function, we set $\mu f = \int f d\mu$ and $Rf : E \to \mathbb{R}$ will be the function defined by $Rf(x) = \int R(x, dy) f(y)$, $x \in E$, provided these integrals are well defined. Finally, the Dirac measure at point $x \in E$ is denoted by δ_x.

2.1. *Contraction and approximation result for the total variation distance.* We remind the reader that the total variation distance between two probability measures $\mu, \nu \in \mathcal{P}(E)$ is defined by

$$\|\mu - \nu\|_{TV} = \sup_{A \in \mathcal{B}(E)} |\mu(A) - \nu(A)| = \frac{1}{2} \sup_{\|f\|_\infty \leq 1} \left| \int f d\mu - \int f d\nu \right|,$$

where for a measurable function $f : E \to \mathbb{R}$, $\|f\|_\infty = \sup_{x \in E} |f(x)|$.

For the family $\{Q_u : u \in [0,1]\}$, the following assumptions will be needed.

A1 There exists an integer $m \geq 1$ and $r \in (0,1)$ such that for all $(u, x, y) \in [0,1] \times E^2$,
$$\|\delta_x Q_u^m - \delta_y Q_u^m\|_{TV} \leq r.$$

A2 There exists a positive real number L such that for all $(u, v, x) \in [0,1]^2 \times E$,
$$\|\delta_x Q_u - \delta_x Q_v\|_{TV} \leq L|u - v|.$$

The Dobrushin contraction coefficient of a Markov kernel R is defined by

$$c(R) = \sup_{\mu \neq \nu \in \mathcal{P}(E)} \frac{\|\mu R - \nu R\|_{TV}}{\|\mu - \nu\|_{TV}} = \sup_{(x,y) \in E^2} \|\delta_x R - \delta_y R\|_{TV}.$$

See for instance Bartoli and Del Moral [4], Theorem 4.3.3, for a proof of the second equality. Note that $c(R) \in [0,1]$. Hence, assumption **A1** means that $\sup_{u \in [0,1]} c(Q_u^m) < 1$. We will still denote by $\| \cdot \|_{TV}$ the total variation distance (or the total variation norm if we consider the space of signed measures) on $\mathcal{P}(E^j)$ for any integer j. Moreover, let $(X_k(u))_{k \in \mathbb{Z}}$ be a stationary Markov chain with transition kernels Q_u, for $u \in [0,1]$. We remind the reader that for an integer $j \geq 1$, $\pi_{k,j}^{(n)}$ (resp. $\pi_{u,j}$) denotes the probability distribution of the vector $(X_{n,k}, \ldots, X_{n,k+j-1})$ (of the vector $(X_k(u), \ldots, X_{k+j-1}(u))$ resp.). The following result is proved in the supplementary material, Section 1.

THEOREM 1. *Suppose that Assumptions* **A1** $-$ **A2** *hold. Then for all* $u \in [0,1]$, *the Markov kernel* Q_u *has a unique invariant probability measure* π_u. *The triangular array of Markov chains* $\{X_{n,k}, n \in \mathbb{Z}^+, k \leq n\}$ *is locally stationary for the total variation distance. Moreover, all integers* $j \geq 1$, *there exists a positive real number* C_j, *not dependent on* k, n, u *and such that*

$$(3) \qquad \|\pi_{k,j}^{(n)} - \pi_{u,j}\|_{TV} \leq C_j \left[\left| u - \frac{k}{n} \right| + \frac{1}{n} \right].$$

Notes.

1. Assumption **A1** is satisfied if there exist a positive real number ε, a positive integer m and a family of probability measures $\{\nu_u : u \in [0,1]\}$ such that

   $$Q_u^m(x, A) \geq \varepsilon \nu_u(A), \text{ for all } (u, x, A) \in [0,1] \times E \times \mathcal{B}(E).$$

   In the homogeneous case, this condition is the so-called Doeblin's condition (see Meyn and Tweedie [36], Chapter 16 for a discussion about

this condition). Note that the lower bound is uniform with respect to x and Doeblin's condition is then mostly interesting when E is compact. To show that this condition is sufficient for **A1**, one can use the inequalities

$$Q_u^m(x,A) - Q_u^m(y,A) \le 1 - \varepsilon + \varepsilon\nu_u(E \setminus A) - Q_u^m(x, E \setminus A) \le 1 - \varepsilon.$$

For a Markov chain with a finite state space, the Doeblin's condition is satisfied if $\min_{(x,y)\in E^2} \inf_{u\in[0,1]} Q_u^m(x,y) > 0$, with ν_u the counting measure on E.

2. One can also consider more general state spaces E. Set $Q_u(x,dy) = f(u,x,y)\mu(x,dy)$ where μ is a probability kernel and the family of conditional densities $\{(x,y) \mapsto f(u,x,y), u \in [0,1]\}$ satisfies

$$\inf_{(u,x,y)\in[0,1]\times E^2} f(u,x,y) \ge \varepsilon, \quad \sup_{x,y\in E} |f(u,x,y) - f(v,x,y)| \le C|u-v|,$$

for positive constants ε and C. In this case, setting $L = C/2$ for **A2**, we have

$$\|\delta_x Q_u - \delta_x Q_v\|_{TV} = \frac{1}{2}\int |f(u,x,y) - f(v,x,y)|\, \mu(x,dy) \le L|u-v|.$$

3. One can also consider higher-order Markov processes, in particular higher-order finite-state Markov chains. For instance if
$\{S(u,\mathbf{x},d\mathbf{y}) = f(u,\mathbf{x},y_m)\gamma(\mathbf{x},dy_m), u \in [0,1]\}$ is a family of probability kernels from $(E^m, \mathcal{E}^{\otimes m})$ to $(E, \mathcal{E})$, we define Q_u as the probability kernel from $(E^m, \mathcal{E}^{\otimes m})$ to itself such that

$$Q_u(\mathbf{x},d\mathbf{y}) = f(u,\mathbf{x},y_m)\mu(\mathbf{x},d\mathbf{y}), \quad \mu(\mathbf{x},d\mathbf{y}) = \prod_{i=1}^{m-1} \delta_{x_{i+1}}(dy_i)\gamma(\mathbf{x},dy_m).$$

If f is lower bounded, then the Doeblin condition is satisfied for Q_u^m. See model (4) discussed below for a particular example of this type. Note that this approach allows to define a nonparametric model in time and space.

2.2. *Uniform mixing properties.* In this subsection, we consider the problem of mixing for the locally stationary Markov chains introduced previously. Such mixing conditions will be crucial to control the limiting behavior of partial sums of locally stationary Markov chains. Under our contraction assumptions (see Assumption **A1**), the stationary Markov chains with Markov

kernels Q_u are uniformly geometric ergodic and the following $\phi-$mixing co-efficients are adapted to our purpose. If $1 \leq i \leq j \leq n$, we set

$$\mathcal{F}_{i,j}^{(n)} = \sigma\left(X_{n,\ell} : i \leq \ell \leq j\right).$$

Now we set for $0 \leq j \leq n-1$,

$$\phi_n(j) = \max_{1 \leq i \leq n-j} \sup \left\{ |\mathbb{P}\left(B|A\right) - \mathbb{P}(B)| : B \in \mathcal{F}_{i+j,n}^{(n)}, A \in \mathcal{F}_{1,i}^{(n)}, \mathbb{P}(A) \neq 0 \right\}.$$

We will say that the triangular array is $\phi-$mixing (or uniformly mixing) if $\phi(j) = \sup_{n \geq j+1} \phi_n(j) \rightarrow_{j \to \infty} 0$. For a time-homogeneous Markov chain, $\phi-$mixing is equivalent to uniform ergodicity (see Ibragimov and Linnik [32], p. 368). Under our assumptions, the $\phi-$mixing coefficients decrease exponentially fast. A proof of the following result is given in the supplementary material, Section 3.1.

PROPOSITION 1. *Assume that assumptions* $\mathbf{A1} - \mathbf{A2}$ *hold true. Then there exist* $C > 0$ *and* $\rho \in (0,1)$, *only depending on* m, L *and* r *such that*

$$\phi(j) \leq C\rho^j.$$

2.3. *Finite state space Markov chains.* This part is devoted to the example of finite-state Markov chains which was our main motivation for this paper. Let E be a finite set. In this case, we obtain the following result. Its proof can be found in the supplementary material, Section 3.2.

COROLLARY 1. *Let* $\{Q_u : u \in [0,1]\}$ *be a family of transition matrices such that for each* $u \in [0,1]$, *the Markov chain with transition matrix* Q_u *is irreducible and aperiodic. Assume further that for all* $(x,y) \in E^2$, *the application* $u \to Q_u(x,y)$ *is Lipschitz continuous. Then Theorem 1 applies and the* $\phi-$mixing *coefficients are bounded as given in Proposition 1.*$\square$

2.4. *Inference of finite-state Markov chains.* We consider the nonparametric kernel estimation of the invariant probability π_u or the transition matrix Q_u. To this end, a classical method used for locally stationary time series is based on kernel estimation. See for instance Dahlhaus and Subba Rao [13], Fryzlewicz et al. [26], Vogt [50] or Zhang and Wu [52] for the nonparametric kernel estimation of locally stationary processes. Let $K : \mathbb{R} \to \mathbb{R}_+$ be a probability density, supported on $[-1,1]$ and of bounded variation. For $b = b_n \in (0,1)$ and $K_b(\cdot) = b^{-1}K(\cdot/b)$, we set

$$e_i(u) = K_b(u - i/n)/\sum_{j=\ell}^{n} K_b(u - j/n), \quad u \in [0,1], \quad \ell \leq i \leq n.$$

Now, we consider some estimators of π_u and Q_u. Let

$$\hat{\pi}_u(x) = \sum_{i=2}^n e_i(u) \mathbb{1}_{\{X_{n,i-1}=x\}} \text{ and } \hat{Q}_u(x,y) = \frac{\hat{\pi}_{u,2}(x,y)}{\hat{\pi}_u(x)},$$

where $\hat{\pi}_{u,2}(x,y) = \sum_{i=2}^n e_i(u) \mathbb{1}_{\{X_{n,i-1}=x, X_{n,i}=y\}}$. Properties of these estimators, which have standard nonparametric rates of convergence, are given in the supplementary material, Section 3.3.

Notes.

1. The estimator $\hat{Q}_u$ is a localized version of the standard estimator used in the homogeneous case. Let us also mention that our estimators coincide with the localized maximum likelihood estimator. Indeed the localized log-likelihood function is defined by

$$\ell_n(P) = \sum_{k=2}^n K_b(u - k/n) \sum_{(x,y)\in E^2} \mathbb{1}_{\{X_{n,k-1}=x, X_{n,k}=y\}} \log\left(P(x,y)\right).$$

   Maximizing this contrast with respect to P under the constraints $\sum_{y\in E} P(x,y) = 1$, $x \in E$, we find a unique solution which is $\hat{Q}_u$. In the homogeneous case, this result was derived in Billingsley [5]. We will give a general result for local maximum likelihood estimators in Section 4.6 in a more general setup.

2. Assuming that $u \mapsto Q_u$ is of class $\mathcal{C}^3$ (i.e. three times continuously differentiable), on can get a second order approximation of the bias $\mathbb{E}\hat{\pi}_u(x) - \pi_u(x)$, $u \in (0,1)$. Here we assume that the kernel K is symmetric. Using the perturbation result given in Cao [11], Section 2, one can show that $u \mapsto \pi_u(x)$ is also of class $\mathcal{C}^3$. Then, using Theorem 1 and some standard properties for the kernel K and setting $b_n(x) = \mathbb{E}\hat{\pi}_u(x) - \pi_u(x)$ and $\kappa_j = \int v^j K(v) dv$, we get

$$\begin{aligned} b_n(x) &= \sum_{i=2}^n e_i(u)\left[\pi_{i/n}(x) - \pi_u(x)\right] + O(1/n) \\ &= -b\pi_u'(x)\kappa_1 + \frac{b^2 \pi_u''(x)}{2}\kappa_2 + O(b^3 + 1/n). \end{aligned}$$

3. One can also study higher-order Markov chains. In this case, vectors of some successive coordinates form a Markov chain of order one and one can apply Theorem 1. However, the number of transitions increases exponentially fast with the order of the Markov chain. A solution for getting parsimonious models is to consider the time-varying versions of the probit or logit models as in Fokianos and

Moysiadis [23] or Moysiadis and Fokianos [37]. For instance, for binary time series taking values 0 or 1, one can consider triangular arrays $\{Z_{n,k} : 1 \le k \le n, n \ge 1\}$ of binary random variables such that

$$(4) \quad \mathbb{P}\left(Z_{n,k} = 1 | Z_{n,k-j}, j \le 1\right) = F\left[a_0(k/n) + \sum_{j=1}^{p} a_j(k/n)Z_{n,k-j}\right],$$

where F is a Lipschitz cumulative distribution function taking values in $(0,1)$ and the a_j's are Lipschitz continuous functions. Local stationarity and mixing properties can be obtained for $X_{n,k} = (Z_{n,k}, \ldots, Z_{n,k-p+1})$ which defines a time-inhomogeneous Markov chain satisfying the assumptions of Theorem 1. Moreover, statistical inference in such models can be conducted using local likelihood estimation as in Dahlhaus and Subba Rao [13] for time-varying ARCH models or Dahlhaus et al. [14]. See Section for a discussion of local maximum likelihood estimation.

3. Local stationarity in Wasserstein metrics. In this section, we consider another metric for considering additional locally stationary Markov chains. This part is more illustrative and only of theoretical interest in this paper. Our aim is to show that many locally stationary autoregressive processes introduced in the literature can be seen as particular examples of locally stationary Markov chains in the sense of Definition 1, using Wasserstein metrics. Many examples are given in the supplementary material (see Section 14) which also contain a discussion about the mixing properties of such triangular arrays (see Section 5). However, the statistical inference will not be developed for this part. In this sequel, we consider a Polish space (E, d). For $p \ge 1$, we consider the set $\mathcal{P}_p(E)$ of probability measures μ on (E, d) such that $\int d(x, x_0)^p \mu(dx) < \infty$. Here x_0 is an arbitrary point in E. It is easily seen that the set $\mathcal{P}_p(E)$ does not depend on x_0. The Wasserstein metric W_p of order p associated with the metric d is defined by

$$(5) \qquad W_p(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \left\{\int_{E \times E} d(x, y)^p d\gamma(x, y)\right\}^{1/p}$$

where $\Gamma(\mu, \nu)$ denotes the collection of all probability measures on $E \times E$ with marginals μ and ν. It is well-known that an optimal coupling always exists when the state space is Polish. An optimal coupling is a coupling γ which realizes the infimum in (5). See Villani [49] for the properties of Wasserstein metrics and the existence of optimal couplings. Note that the total variation distance can be seen as a particular example of the Wasserstein metric by setting $d(x, y) = \mathbb{1}_{x \ne y}$ and the case of finite-state Markov chains can be

treated with the results of this section. However, in this section, we have in mind the case of an Euclidean norm on $E = \mathbb{R}^p$ or more generally a Banach space $(E, \| \cdot \|)$. In this case, Dedecker and Prieur [16] have defined some mixing coefficients for contracting Markov chains. But we will keep a general metric d for the exposure. Note that for a real-valued Markov chain defined by (2), we have $Q_u(x, A) = \mathbb{P}\left(F_{\theta(u)}(x, \varepsilon_1) \in A\right)$ and

$$W_p\left(\delta_x Q_u, \delta_y Q_v\right) \leq \|F_{\theta(u)}(x, \varepsilon_1) - F_{\theta(u)}(y, \varepsilon_1)\|_p.$$

This inequality shows that an autoregressive process contracting for the $\mathbb{L}^p$−norm (i.e the right hand term of the previous equation can be bounded by $\alpha|x-y|$ with $\alpha \in (0, 1)$) entails a contraction of the corresponding Markov kernel in Wasserstein metric. This will be a particular case of our general result.

In the sequel, we will use the following assumptions.

B1 For all $(u, x) \in [0, 1] \times E$, $\delta_x Q_u \in \mathcal{P}_p(E)$.

B2 There exist a positive integer m and two real numbers $r \in (0, 1)$ and $C_1 \geq 1$ such that for all $u \in [0, 1]$ and all $x \in E$,

$$W_p\left(\delta_x Q_u^m, \delta_y Q_u^m\right) \leq rd(x, y), \quad W_p\left(\delta_x Q_u, \delta_y Q_u\right) \leq C_1 d(x, y).$$

B3 The family of Markov kernels $\{Q_u : u \in [0, 1]\}$ satisfies the following Lipschitz continuity condition. There exists $C_2 > 0$ such that for all $x \in E$ and all $u, v \in [0, 1]$,

$$W_p\left(\delta_x Q_u, \delta_x Q_v\right) \leq C_2\left(1 + d(x, x_0)\right)|u - v|.$$

Note. If R is a Markov kernel, the Dobrushin contraction coefficient is now defined by $c(R) := \sup_{\mu \neq \nu \in \mathcal{P}_p(E)} \frac{W_p(\mu R, \nu R)}{W_p(\mu, \nu)} = \sup_{\substack{(x,y) \in E^2 \\ x \neq y}} \frac{W_p(\delta_x R, \delta_y R)}{d(x,y)}$. The second inequality is guaranteed using Lemma 5 (2.) given in the supplementary material and the equality $W_p(\delta_x, \delta_y) = d(x, y)$. Then Assumption **B2** means that $\sup_{u \in [0,1]} c(Q_u) < \infty$ and $\sup_{u \in [0,1]} c(Q_u^m) < 1$. Using straightforward arguments, one can check that under Assumptions **B1-B3**, a chain with transition kernel Q_u is geometrically ergodic and its unique invariant probability measure π_u is Lipschitz continuous with respect to u, for the metric W_p. See Section 5 in the supplementary material.

Now let us present the main result of this section. For $j \in \mathbb{N}^*$, we endow the space E^j with the distance $d_j(x, y) = (\sum_{s=1}^{j} d(x_s, y_s)^p)^{1/p}, \quad x, y \in E^j$. We will still denote by W_p the Wasserstein metric for Borel measures on E^j. The proof of the following result can be found in the supplementary material, Section 4.

THEOREM 2. *Assume that assumptions $\mathbf{B1} - \mathbf{B3}$ hold true. Then the triangular array of Markov chains $\{X_{n,k} : n \in \mathbb{Z}^+, k \leq n\}$ is locally stationary. For all integers j, there exists a real number $C > 0$ such that for all $u \in [0,1]$ and $1 \leq k \leq n - j + 1$,*

$$W_p\left(\pi_{k,j}^{(n)}, \pi_{u,j}\right) \leq C\left[\left|u - \frac{k}{n}\right| + \frac{1}{n}\right].$$

Notes.

1. Theorem 2 can be used to approximate some expectations $\int f d\pi_{k,j}^{(n)}$ when $f : E^j \to \mathbb{R}$ is a smooth function. For instance, assume that there exist $C > 0$ such that for all $(z_1, z_2) \in E^j \times E^j$,

$$|f(z_1) - f(z_2)| \leq C\left[1 + d_j(z_1, x_0)^{p-1} + d_j(z_2, x_0)^{p-1}\right] \cdot d_j(z_1, z_2)$$

and let $\gamma_{k,u}^{(n)}$ be a coupling of $\left(\pi_{k,j}^{(n)}, \pi_{u,j}\right)$. Using the Hölder inequality, we have, setting $q = \frac{p}{p-1}$,

$$\left|\int f d\pi_{k,j}^{(n)} - \int f d\pi_{u,j}\right| \leq \int |f(z_1) - f(z_2)| \, d\gamma_{k,n,u}(z_1, z_2)$$

$$\leq C3^{\frac{q-1}{q}} d_{k,u}^{(n)} \left(\int d_j(z_1, z_2)^p d\gamma_{k,u}^{(n)}(z_1, z_2)\right)^{1/p},$$

$$d_{k,u}^{(n)} = 1 + \left(\int d_j(z_1, x_0)^p d\pi_{k,j}^{(n)}(z_1)\right)^{1/q} + \left(\int d_j(z_1, x_0)^p d\pi_{u,j}(z_1)\right)^{1/q}.$$

Then under the assumptions of Theorem 2, there exists a constant $D > 0$ such that

$$\left|\int f d\pi_{k,j}^{(n)} - \int f d\pi_{u,j}\right| \leq D\left[\left|u - \frac{k}{n}\right| + \frac{1}{n}\right].$$

2. When E is a separable Banach space, and a stationary Markov kernel is contracting for the Wasserstein metric, Dedecker and Prieur [16] have defined the coefficients of $\tau-$dependence coefficients for the corresponding Markov chains. Since we will not use these coefficients in the rest of the paper, we defer the reader to the supplementary material, Section 7, for a discussion of these coefficients. Using these mixing coefficients and the previous point, a study of the asymptotic behavior of localized partial sums $\sum_{k=1}^{n-j} e_i(u)f(X_{n,k}, \ldots, X_{n,k+j-1})$ as in Section 2 or in Section 4 should be possible. For autoregressive processes contracting in $\mathbb{L}^p$, some results such as Theorem 2.10 in Dahlhaus et al. [14] can be also obtained from our approach.

3.1. *Examples of locally stationary Markov chains .*   In this part, we give two examples different from the standard autoregressive processes considered in the literature. Additional examples as well as some justifications can be found in the supplementary material, Section 11.

A locally stationary functional time series.   As suggested by one referee, we give an example of locally stationary functional time series. Stationary functional time series have received a considerable attention over the recent years. See for instance Horváth and Kokoszka [31] for a recent survey. Here, we provide a locally stationary version of a very simple functional autoregressive process. See Horváth and Kokoszka [31], Chapter 13 for a more general stationary version. Let $p = 2$, $E = \mathbb{L}^2([0,1])$ and $d(x,y)^2 = \int_0^1 (x(s) - y(s))^2 \, ds$. Let $B_1, B_2, \ldots$ be a sequence of independent Brownian motions over $[0,1]$. We assume that

$$X_{n,k}(t) = \int_0^1 a_{k/n}(t,s) X_{n,k-1}(s) ds + \int_0^1 \sigma_{k/n}(t,s) dB_k(s), \quad t \in [0,1],$$

where the kernel functions a_u and σ_u satisfy, for a constant $C > 0$,

$$\sup_{u \in [0,1]} \int_0^1 \int_0^1 a_u(t,s)^2 ds dt < 1, \quad \sup_{u \in [0,1]} \int_0^1 \int_0^1 \sigma_u(t,s)^2 ds dt < \infty,$$

$$\int_0^1 \int_0^1 \left[|a_u(t,s) - a_v(t,s)|^2 + |\sigma_u(t,s) - \sigma_v(t,s)|^2 \right] ds dt < C^2 |u - v|^2.$$

Here $\delta_x Q_u$ is defined as the probability distribution of the random variable $\int_0^1 a_u(t,s) x(s) ds + \int_0^1 \sigma_u(t,s) dB_1(s)$. Additional justifications for this example are given in the supplementary material, Section 6.

Poisson GARCH process.   Stationary Poisson GARCH processes are widely used for analyzing series of counts. See Fokianos et al. [24] for the properties and the statistical inference of such processes. In this paper, we consider a time-varying version of this model. More precisely, we assume that the conditional distribution $Y_{n,k} | \sigma (Y_{n,k-j}, j \geq 1)$ is a Poisson distribution of parameter $\lambda_{n,k}$ given recursively by

$$\lambda_{n,k} = \gamma(k/n) + \alpha(k/n) Y_{n,k-1} + \beta(n,k) \lambda_{n,k-1}, \quad \max_{u \in [0,1]} [\alpha(u) + \beta(u)] < 1,$$

where γ, α, β are positive Lipschitz functions such that To construct a Markov chain, we consider $X_{n,k} = (Y_{n,k}, \lambda_{n,k})'$. One can show that our assumptions are satisfied for $p = 1$ and $d(x,y) = \sum_{i=1}^2 |x_i - y_i|$. A coupling of different paths can be obtained using Poisson processes. See the supplementary

material, Section 8, for details. However, let us mention that this result only guarantees the approximation of integral $\int f d\pi_{u,j}$ for Lipschitz functions which seems to be too restrictive for statistical inference. On the other hand, contraction for $p > 1$ is unclear. Let us also mention that approximation in W_p-metric, $p > 1$, seems not satisfying for the Poisson distribution. Indeed, suppose that μ_u denotes the Poisson distribution of parameter λ_u. Since $x^p \geq x$ for any nonnegative integer x, we have

$$W_p\left(\mu_u, \mu_v\right) \geq W_1\left(\mu_u, \mu_v\right)^{1/p} = |\lambda_u - \lambda_v|^{1/p}.$$

Then if $u \mapsto \lambda_u$ is differentiable, we have $W_p\left(\mu_{u+h}, \mu_u\right) \sim h^{1/p}$ in a neighborhood of point u. Then the regularity in W_p-metric is only of Hölder type even for a smooth functional parameter which is an undesirable property. This is why for integer-valued processes, we will use the results of the next section which will give sharper results for the approximation of the stationary distributions. Unfortunately, the assumptions will be only satisfied for the Poisson ARCH process ($\beta = 0$). See Section 4.3 for details.

4. Local stationarity from drift and small set conditions. Our motivation for this section is to define some locally stationary versions of Markov chains models that satisfy a drift and a small set condition. This approach will be interesting for unbounded state spaces and models for which

1. the Doeblin condition discussed in Section 2 does not hold,
2. it is difficult or even impossible to get a natural coupling of the Markov kernels in such a way the contraction and continuity conditions **B2-B3** are satisfied when E is a separable Banach space.

This is in particular the case for some integer-valued autoregressive processes such as Poisson ARCH, for which we already mentioned the difficulties in using the W_p-metric for $p > 1$. But we will also consider additional models such as Markov-switching or SETAR processes for which contraction in Wassertein metric with the euclidean metric seems difficult to get. This section can be seen as an extension of the Dobrushin's contraction technique used in Section 2. A key point for this is a result obtained by Hairer and Mattingly [28] who revisited geometric ergodicity using contraction properties of Markov kernels with respect to some $V-$norms. This result can also be found in Douc et al. [20], Lemma 6.29, a reference in which the authors give many examples of autoregressive processes satisfying the corresponding assumptions. From this important result (see Lemma 6 in the supplementary material for a statement in our context), we will consider additional examples of locally stationary Markov chains with unbounded state spaces.

For a positive real number ϵ and a positive integer m, we set

$$I_m(\epsilon) = \left\{ (u_1, \ldots, u_m) \in [0, 1]^m : |u_i - u_j| \leq \epsilon, \quad 1 \leq i \neq j \leq m \right\}.$$

For a function $V : E \to [1, \infty)$, we define the $V-$norm of a signed measure μ on $\left(E^j, \mathcal{B}(E^j) \right)$ by

$$(6) \qquad \|\mu\|_V = \sup \left\{ \int f d\mu : \quad |f(x_1, \ldots, x_j)| \leq V(x_1) + \cdots + V(x_j) \right\}.$$

4.1. *General result.* Let $V : E \to [1, \infty)$ be a measurable function, ϵ a positive real number and m a positive integer. We will use the following assumptions.

F1 there exist $\lambda \in (0, 1)$ and two real numbers $b > 0, K \geq 1$ such that for all $(u_1, \ldots, u_m) \in I_m(\epsilon)$,

$$Q_{u_1} V \leq KV, \quad Q_{u_1} \cdots Q_{u_m} V \leq \lambda V + b.$$

F2 There exist $\eta > 0$, $R > 2b/(1 - \lambda)$ (λ and b are defined in the previous assumption) and a probability measure $\nu \in \mathcal{P}(E)$ such that for

$$\delta_x Q_{u_1} \cdots Q_{u_m} \geq \eta\nu, \text{ if } V(x) \leq R,$$

F3 there exists a function $\widetilde{V} : E \to (0, \infty)$ such that $\sup_{u \in [0,1]} \pi_u \widetilde{V} < \infty$ and for all $x \in E$, $\|\delta_x Q_u - \delta_x Q_v\|_V \leq \widetilde{V}(x)|u - v|$.

We first give some properties of the Markov kernels Q_u with respect to the $V-$norm. The proof of the next proposition can be found in the supplementary material, Section 12.

PROPOSITION 2. *Suppose that Assumptions* **F1** $-$ **F3** *hold. Then the two following statements are valid.*

1. *There exist $C > 0$ and $\rho \in (0, 1)$ such that for all $x \in E$,*

$$\sup_{u \in [0,1]} \|\delta_x Q_u^j - \pi_u\|_V \leq CV(x)\rho^j, \quad \sup_{u \in [0,1]} \pi_u V < \infty.$$

2. *There exists $C > 0$ such that for all $(u, v) \in [0, 1]^2$, $\|\pi_u - \pi_v\|_V \leq C |u - v|.$*

Now, we give our result about local stationarity. A proof can be found in the supplementary material, Section 13.

Theorem 3.   *1. Suppose that Assumptions **F1** − **F2** hold true. Then there exists a positive real number C, only depending on m, λ, b, K and $\sup_{u \in [0,1]} \pi_u \widetilde{V}$ such that*

$$\|\pi_k^{(n)} - \pi_u\|_V \leq C \left[\left| u - \frac{k}{n} \right| + \frac{1}{n} \right].$$

2. In addition, suppose that for all $(u, v) \in [0, 1]^2$,

$$\|\delta_x Q_u - \delta_x Q_v\|_{TV} \leq L(x)|u - v| \quad \text{with} \quad \sup_{\substack{u \in [0,1] \\ 1 \leq \ell' \leq \ell}} \mathbb{E}\left[L\left(X_\ell(u)\right) V\left(X_{\ell'}(u)\right) \right] < \infty.$$

Let $j \geq 1$ be an integer. Then there exists $C_j > 0$, not depending on k, n, u and such that

$$(7) \qquad \|\pi_{k,j}^{(n)} - \pi_{u,j}\|_V \leq C_j \left[\left| u - \frac{k}{n} \right| + \frac{1}{n} \right].$$

Moreover, the triangular array of Markov chains $\{X_{n,k} : n \in \mathbb{Z}^+, k \leq n\}$ is locally stationary.

Notes.

1. The continuity assumption **F3** means that the application $u \mapsto Q_u$ is Lipschitz continuous for a particular operator norm. More precisely, we set

$$\|Q_u\|_{V, \widetilde{V}} := \sup_{\|\mu\|_{\widetilde{V}} \leq 1} \|\mu Q_u\|_V = \sup_{|f|_V \leq 1} |Q_u f|_{\widetilde{V}},$$

   where $|f|_V = \sup_{x \in E} \frac{|f(x)|}{V(x)}$. The equality between the two expressions given above results from straightforward computations. Assumption **F3** is then equivalent to the continuity of the application $u \mapsto Q_u$ for the norm $\| \cdot \|_{V, \widetilde{V}}$. An important remark is the following. For unbounded state spaces, the continuity of the Markov kernel may not hold if $V = \widetilde{V}$. For instance Ferré et al. [22] have shown that for an AR(1) and $V(x) = 1 + |x|$, the transition kernel is never continuous for this simpler norm, whatever the density of the absolutely continuous noise distribution. In general, it is necessary to choose a function $\widetilde{V}$ larger than V to get the continuity of the Markov kernels with respect to the parameters of the models.

2. Note that this result automatically gives the bound

$$\mathbb{P}\left(X_{n,k} \in A\right) - \mathbb{P}\left(X_k(u) \in A\right) = O\left(|u - k/n| + 1/n\right)$$

   for any measurable set A.

3. In the stationary case, conditions **F1** and **F2** are often satisfied for autoregressive processes for which the regression function is contracting outside a ball. Locally stationary versions of autoregressive processes satisfying this contraction "at infinity" have been studied by Vogt [50]. We claim that his conditions more or less guarantee our assumptions when V is a suitable power function, because his results are based on assumptions ensuring geometric ergodicity in the stationary case. For stationary autoregressive models, we refer the reader to Douc et al. [20], Chapter 6 for some examples satisfying Assumptions **F1-F2**.

Assumptions **F1-F2** also guarantee some mixing properties for the triangular array of Markov chains. In our context, the notion of $\beta-$mixing is adapted. See Doukhan [21] for the various mixing notions adapted to random sequences and in particular the properties of $\beta-$mixing and $\phi-$mixing sequences. In particular, the $\phi-$mixing property derived in Section 2 implies the $\beta-$mixing property discussed in the present section. However, for Markov chains, the $\phi-$mixing property is related to the Doeblin's condition and rarely holds for unbounded state spaces. See Bradley et al. [9] for a discussion of different mixing conditions for Markov chains. For Markov chains, the $\beta-$mixing coefficients can be defined as follows (see Proposition 3.22 in Bradley [8]). For an integer $n \geq 1$ and $0 \leq j \leq n$, we set

$$\beta_n(j) = \frac{1}{2} \max_{1 \leq i \leq n-j} \mathbb{E} \sup_{\|f\|_\infty \leq 1} |\mathbb{E}(f(X_{n,i+j})|X_{n,i}) - \mathbb{E}f(X_{n,i+j})| .$$

Similarly, we define the coefficients $(\beta^{(u)}(j))_{j \geq 0}$ of the stationary Markov chain $(X_k(u))_{k \in \mathbb{Z}}$. The following result, whose proof is straightforward, can be obtained by bounding the larger coefficients

$$\beta_n^{(V)}(j) = \sup_{k \leq n} \mathbb{E}\|\pi_k^{(n)} - \delta_{X_{n,k-j}} Q_{\frac{k-j+1}{n}} \cdots Q_{\frac{k}{n}}\|_V.$$

PROPOSITION 3. *Assume that Assumptions $\mathbf{F1} - \mathbf{F2}$ hold true and that $n \geq m/\epsilon$. Then if $j = mg + s$, we have*

$$(8) \quad \beta_n(j) \leq \delta^{-1} \sup_{k \leq n} \pi_k^{(n)} V \cdot K^s \gamma^g, \quad \beta^{(u)}(j) \leq \delta^{-1} \sup_{u \in [0,1]} \pi_u V \cdot K^s \gamma^g$$

where $\delta, \gamma \in (0,1)$ are given in Lemma 6 of the supplementary material.

Notes.

1. In checking assumptions **F1-F2**, one can find some conditions under which the time-varying ARCH process is $\beta-$mixing. This gives a short

alternative proof to the result derived in Fryzlewicz and Subba Rao [25]. See the supplementary material, Section 14 for precise assumptions.

2. From the drift condition in **F1**, we have $\pi_k^{(n)} V \leq \frac{b}{1-\lambda}$ for $n \geq m/\epsilon$. Hence $\sup_{n \in \mathbb{Z}^+, k \leq n} \pi_k^{(n)} V < \infty$.

3. Assumptions $\mathbf{F1} - \mathbf{F2}$ guarantee a geometric decay for the $\beta-$mixing coefficients of the triangular array of Markov chains and of the corresponding stationary approximations. One can observe that the bound can be made uniform over $n \geq 1$ and $u \in [0, 1]$.

4.2. *Example 1: Markov switching autoregressive processes.* Markov switching autoregressive processes have been introduced by Hamilton [30] to analyze business cycles in economics. These processes have been widely studied in the literature. We consider a simple autoregressive process with one lag and regime switching and which is a locally stationary version of the CHARME model considered in Stockis et al. [44]. Let $\{Z_{n,k} : 1 \leq k \leq n, n \geq 1\}$ be a triangular array of Markov chains on a finite state space E_2 and associated with a family $\{\overline{Q}_u : u \in [0, 1]\}$ of transition matrices which are assumed to satisfy the assumptions of Corollary 1. We also consider a sequence $(\varepsilon_n)_{n \in \mathbb{Z}}$ of i.i.d. random variables with an absolutely continuous distribution. Then we define

$$Y_{n,k} = m\left(\frac{k}{n}, Z_{n,k}, Y_{n,k-1}\right) + \sigma\left(\frac{k}{n}, Z_{n,k}, Y_{n,k-1}\right)\varepsilon_k, \quad 1 \leq k \leq n,$$

where $m : [0,1] \times E_2 \times \mathbb{R} \to \mathbb{R}$ and $\sigma : [0,1] \times E_2 \times \mathbb{R} \to \mathbb{R}_+^*$ are given functions. We assume that for each $n \geq 1$, the Markov chain $(Z_{n,k})_k$ is independent of the sequence $(\varepsilon_k)_k$. We set $E = \mathbb{R} \times E_2$ and $X_{n,k} = (Y_{n,k}, Z_{n,k})'$. Then as for the homogeneous case, the bivariate process $(X_{n,k})_k$ forms a Markov chain. One can choose some power drift functions $V(y, z) = 1 + |y|^p$. More precise assumptions on the noise density f_ε and the functions m, σ that guaranty local stationarity can be found in the supplementary material, Section 15. When the functions m, σ do not depend on the time, weaker assumptions can be used. See the supplementary material for details.

4.3. *Example 2: Integer-valued autoregressive processes .* Stationary INAR processes are widely used in the time series community for analyzing integer-valued data. This time series model has been proposed by Al Osh and Alzaid [1] and a generalization to several lags was studied in Jin-Guan and Yuan [33]. In this paper, we introduce a locally stationary version of such processes, with one lag for simplicity. For $u \in [0, 1]$, we consider a random binomial operator $\alpha_u \circ$, i.e. for each integer x, $\alpha_u \circ x$ follows a binomial distribution with

parameters (x, α_u). One can also set $\alpha_u \circ x = \sum_{i=1}^x Y_i(u)$ where $(Y_i(u))_{i \geq 1}$ is a sequence of i.i.d. Bernoulli random variables. Moreover, let $\zeta(u)$ be the Poisson distribution on the nonnegative integers with mean $\lambda(u)$. Now let

$$X_{n,k} = \alpha_{k/n} \circ X_{n,k-1} + \eta_{n,k}, \quad 2 \leq k \leq n$$

where for each integer $n \geq 1$, $\eta_{n,k}$, which is assumed to be independent of $\alpha_{k/n}\circ$, has probability distribution $\zeta(k/n)$. Note that if the Bernoulli random variables are replaced with Poisson random variables, we obtain the Poisson ARCH process already discussed in the last section. In both case, the parameters α_u (λ_u resp.) of the counting sequence (the Poisson noise resp.) are assumed to be Lipschitz continuous and $\max_{u \in [0,1]} \alpha_u < 1$. One can then show that Theorem 3 applies with drift functions $V_p(x) = x^p + 1$ for an arbitrary integer $p \geq 1$. Details are given in the supplementary material, Section 16. Parameters α_u and λ_u can be estimated using a local least squares method. Details are given in the supplementary material, Section 17. Note that, as in the homogeneous case, the Poisson GARCH process does not satisfies a small set condition. See Fokianos et al. [24] for a discussion.

4.4. *Example* 3: *a locally stationary version of SETAR processes.* We assume here that

$$
\begin{aligned}
X_{n,k} &= (a(k/n)X_{n,k-1} + b(k/n)) \, \mathbb{1}_{\{X_{n,k-1} < r\}} \\
&+ (c(k/n)X_{n,k-1} + d(k/n)) \, \mathbb{1}_{\{X_{n,k-1} \geq r\}} + \varepsilon_k,
\end{aligned}
$$

where $(\varepsilon_k)_{k \in \mathbb{Z}}$ is a sequence of i.i.d. random variables with mean zero. This model is a time-varying version of the stationary threshold model of Tong [46]. Note that the threshold level r is not time-varying, otherwise Assumptions **F3** cannot be checked. We assume that the functions a, b, c, d are Lipschitz continuous with $\alpha = \max(\max_{u \in [0,1]} |a(u)|, \max_{u \in [0,1]} |c(u)|) < 1$ and the noise has a density f_ε of class $\mathcal{C}^1$, positive everywhere and such that for a positive integer $p > 0$,

$$\int |z|^{p+1} f_\varepsilon(z) dz < \infty, \quad \int |z|^p \cdot \left| f_\varepsilon'(z) \right| dz < \infty.$$

Then local stationarity in $V-$norm holds, with $V(y) = 1 + |y|^p$. Justifications are given in the supplementary material, Section 18. If the threshold parameter r is known, local least-squares estimators for a, b, c, d can be shown to be asymptotically Gaussian, using Theorem 3 and Proposition 4 below. Details are omitted. Estimating r is more difficult. One solution could be to estimate it in a second step after plugging the estimates of the autoregressive parameters. See for instance Li et al. [35], Section 3.1 for details in the stationary case.

4.5. *Limiting behavior of partial sums.* In this subsection, we show how our results can be used to obtain some asymptotic normality properties for partial sums. We will use the following terminology. The triangular array of Markov chains $\{X_{n,k} : 1 \leq k \leq n, n \geq 1\}$ will be said to be locally stationary for the $V-$norm if (7) holds and geometrically $\beta-$mixing if (8) holds. Conditions ensuring both properties are given in Theorem 3. The following result is central to derive asymptotic properties of local least-squares or local likelihood estimators. Its proof uses a central limit theorem for strong mixing sequences and is given in the supplementary material, Section 19.

PROPOSITION 4. *Let $\{X_{n,k} : 1 \leq k \leq n, n \geq 1\}$ be a triangular array of Markov chains locally stationary for the $V-$norm and geometrically $\beta-$mixing. For some integer $j \geq 1$, let $f : [0,1] \times E^j \to \mathbb{R}$ be a measurable function continuous with respect to its first argument and $\delta \in (0,1)$ such that*

$$\sup_{u \in [0,1], x_1, \ldots, x_j \in E} \frac{|f(u, x_1, \ldots, x_j)|}{[V(x_1) + \cdots + V(x_j)]^{\frac{1}{2+\delta}}} < \infty.$$

For $j \leq k \leq n$, we set $Z_{n,k} = f(k/n, X_{n,k-j+1}, \ldots, X_{n,k})$ and $Z_k(u) = f(u, X_{k-j+1}(u), \ldots, X_k(u))$ for the stationary approximation. We have the two following properties.

1. *The partial sum $S_n := \frac{1}{\sqrt{n}} \sum_{i=j}^{n} [Z_{n,i} - \mathbb{E}Z_{n,i}]$ is asymptotically Gaussian with mean 0 and variance*

$$\sigma^2 = \int_0^1 \sum_{k \in \mathbb{Z}} Cov(Z_0(u), Z_k(u)) \, du.$$

2. *If K is a kernel of bounded variation and with compact support $[-1,1]$ and $b = b_n$ is such that $b \to 0$ and $nb \to \infty$, then the weighted partial sum $S_n(u) := \frac{1}{\sqrt{nb}} \sum_{i=j}^{n} K(\frac{u-i/n}{b}) [Z_{n,i} - \mathbb{E}Z_{n,i}]$ is asymptotically Gaussian with mean 0 and variance*

$$\sigma(u)^2 = \int_{-1}^{1} K^2(v) dv \cdot \sum_{k \in \mathbb{Z}} Cov(Z_0(u), Z_k(u)).$$

4.6. *Statistical inference of local parameters .* For models satisfying the assumptions of Theorem 3, we derive the asymptotic properties of the local maximum likelihood estimator in the spirit of the recent approach used in Dahlhaus et al. [14] for autoregressive processes. We assume that the family of Markov kernels $\{Q_u : u \in [0,1]\}$ satisfies the assumptions $\mathbf{F1} - \mathbf{F3}$ and that

$$Q_u(x, dy) = exp(S(\theta_0(u), x, y)) \mu(x, dy)$$

where $\theta_0 : [0,1] \mapsto \Theta$ is a function taking values in a subset Θ of $\mathbb{R}^d$, $S : \Theta \times E^2 \to \mathbb{R}$ is a known function and μ is a measure kernel from $(E, \mathcal{E})$ to itself. The local MLE at point $u \in (0,1)$ is defined by

$$\hat{\theta}(u) = \arg\max_{\theta \in \Theta} \mathcal{L}_n(\theta), \quad \mathcal{L}_n(\theta) = n^{-1} \sum_{j=2}^{n} K_b(u - j/n) S\left(\theta, X_{n,j-1}, X_{n,j}\right).$$

Let $\nabla_1 f$, $\nabla_1^2 f$ be the gradient vector and the Hessian matrix with respect to the first argument of a real-valued function f. The following assumptions will be needed.

L1(ℓ) For all $(x, y) \in E^2$, the function $\theta \mapsto S(\theta, x, y)$ is of class $\mathcal{C}^\ell$.

L2 There exist a constant $C > 0$ such that for all $(x, y) \in E^2$,

$$\sup_{\theta \in \Theta} |S(\theta, x, y)|^{2+\delta} \leq C\left[V(x) + V(y)\right].$$

L3 Θ is a compact set and $\theta_0(u) \in int(\Theta)$ is the unique minimizer of $\theta \mapsto \mathbb{E}S(\theta, X_0(u), X_1(u))$ over $\theta \in \Theta$.

L4 There exist a constant $C > 0$ such that for all $(x, y) \in E^2$,

$$\sup_{\theta \in \Theta} \left[|\nabla_1 S(\theta, x, y)|^{2+\delta} + \left|\nabla_1^2 S(\theta, x, y)\right|^{2+\delta} \right] \leq C\left[V(x) + V(y)\right].$$

L5 The function $g_u : v \mapsto \mathbb{E}\nabla_1 S(\theta_0(u), X_{k-1}(v), X_k(v))$ is of class $\mathcal{C}^2$.

Assumptions **L2** and **L4** (for the second derivative) are probably not optimal because we did not prove a sharp law of large number for localized sums. However they are sufficient for illustrating our results. Here we propose to expand the bias of the local MLE up to the second order. A proof of the following result can be found in the supplementary material, Section 20.

THEOREM 4. *Let K be a symmetric kernel, supported on $[-1,1]$ and of bounded variation.*

1. *Let $b \to 0$ and $nb \to \infty$. If Assumptions **L1(0)** and $L2 - L3$ hold true, then $\hat{\theta}(u)$ is consistent.*

2. *If $b \to 0$, $nb \to \infty$, $nb^5 = O(1)$ and the assumptions **L1(2)** and **L2-L5** hold true, then*

$$\sqrt{nb}\left(\hat{\theta}(u) - \theta_0(u) - \frac{b^2 \kappa_2}{2}\zeta(u)\right) \Rightarrow \mathcal{N}\left(0, \int K^2(v)dv \cdot M(u)^{-1}\right),$$

with $M(u) = \mathbb{E}\left[-\nabla_1^2 S(\theta_0(u), X_0(u), X_1(u))\right]$, $\zeta(u) = M(u)^{-1}g_u''(u)$ and $\kappa_2 = \int v^2 K(v)dv$.

Notes.

1. In Section 20 of the supplementary material, we check the previous assumptions for the binary time series (4) and the Poisson ARCH process. Assumption **L5** is not guaranteed by our approximation results. In Dahlhaus et al. [14], an expression for the first order approximation of the bias of the local likelihood estimator is obtained by using a notion of derivative process $\frac{d}{du}X_t(u)$. For categorical or integer-valued data, this notion does not make sense. However, it is still possible to study the regularity of the function $u \mapsto \int f d\pi_{u,2} = \mathbb{E}f\left(X_0(u), X_1(u)\right)$. Using our results, it is only possible to obtain the Lipschitz continuity of this function. However, in Truquet [47], we give a general result which guarantees existence of derivatives for such functions when power functions satisfy the drift condition. In the supplementary material, we use this result to show that Assumption **L5** is satisfied for binary time series and Poisson ARCH process as soon as $u \mapsto \theta(u)$ is two times continuously differentiable.

2. Theorem 4 can also be applied to some standard autoregressive models such as ARCH processes for instance. However, the assumptions **F2-F3** can only be checked using some smoothness assumptions for the noise density. Our assumptions are then more restrictive than that of Dahlhaus et al. [14] which do not require existence of a density for the noise distribution. On the other hand, considering for instance the EXPAR model studied in Dahlhaus et al. [14], Proposition 3 and the result of Section 4.3 in Truquet [47] can be used to check Assumption **L5** under suitable moment conditions on the noise density and its derivatives. An expansion of the bias of the local MLE (or QMLE for non Gaussian inputs) up to any order is also possible. We will not give precise assumptions in the present paper.

5. Practical implementation for finite-state Markov chains. This section is devoted to the implementation of finite-state Markov chains. In particular, we discuss bandwidth selection and prove the consistency of an adapted bootstrap procedure for getting confidence intervals for the elements of the transition matrix.

5.1. *Simulation study.* One of the important issue for the practical implementation of our estimator is bandwidth selection. Interpreting our estimator as a least-squares estimator, we propose a very simple procedure based on Generalized cross validation. The same approach can be used for local least-squares estimators in other locally stationary Markov chain mod-

els such as the time-varying integer valued process discussed in Section 4. For some $y \in E$, we know that

$$\mathbb{P}\left(X_{n,k} = y | X_{n,k-1}\right) = Q_{k/n}\left(X_{n,k-1}, y\right) = \sum_{x \in E} Q_{k/n}(x, y) \mathbb{1}_{\{X_{n,k-1}=x\}}.$$

Moreover, $\hat{Q}_{k/n}$ is a minimizer of the loss function

$$P \mapsto \sum_{y \in E} \sum_{j=2}^{n} K\left(\frac{k-j}{nb}\right) \left(\mathbb{1}_{\{X_{n,j}=y\}} - P\left(X_{n,k-1}, y\right)\right)^2.$$

Then $Z_{y,j} = \mathbb{1}_{\{X_{n,j}=y\}}$ has a fitted value $\hat{Z}_{y,j} = \hat{Q}_{j/n}\left(X_{n,j-1}, y\right)$. The hat matrix H defined by the equality by $\hat{Z} = HZ$ has diagonal elements not depending on y and given by

$$g_j = \sum_{x \in E} \frac{K(0)\mathbb{1}_{\{X_{n,j-1}=x\}}}{\sum_{i=2}^{n} K\left(\frac{k-i}{nb}\right) \mathbb{1}_{\{X_{n,i-1}=x\}}}.$$

Then one can minimize the criterion $\mathcal{C}$ defined by

$$\mathcal{C}(b) = \sum_{y \in E} \frac{\sum_{k=2}^{n} \left(\mathbb{1}_{\{X_{n,j}=y\}} - Q_{k/n}\left(X_{n,k-1}, y\right)\right)^2}{\left(1 - \frac{1}{n-1}\sum_{k=2}^{n} g_k\right)^2}.$$

Next, we perform a simulation study and approximate the mean squared error for estimating the transition matrix for a binary Markov chain under two scenarios.

- For the first scenario $S1$, we set $Q_u(0, 1) = 0.5 + 0.4\sin(2\pi u)$ and $Q_u(1, 0) = 0.5 + 0.4\cos(2\pi u)$.
- For the second scenario $S2$, we set $Q_u(0, 1) = 0.1 + 0.8u$ and $Q_u(1, 0) = 0.9 - 0.8u$.

We evaluate the RMSE $\sqrt{\mathbb{E}\int_0^1 \left(\hat{Q}_u(0, 1) - Q_u(0, 1)\right)^2 du}$ using 5000 samples of size $n = 150$ or $n = 500$. Results are reported in Table 1. With respect to other possible bandwidths, the cross-validation works quite well even for the smallest sample size. For the scenario $S1$ and $n = 150$, the case $b = n^{-1/3}$ is not reported because for some time subscripts t, we have $X_{n,s} \neq 0$ for $s \in [t - nb, t + nb]$. In practice, this problem can be avoided by considering bandwidth parameters large enough so that for each time interval of the previous form, one can find the realization 0 for the process.

TABLE 1
Approximation of the RMSE for several bandwidth choices and sample sizes $n = 150, 500$

	$n = 150$				$n = 500$			
	GCV	$n^{-1/3}$	$1.5n^{-1/3}$	$2n^{-1/3}$	GCV	$n^{-1/3}$	$1.5n^{-1/3}$	$2n^{-1/3}$
$S1$	0.1278	$\times$	0.1340	0.1602	0.076	0.0847	0.0708	0.1103
$S2$	0.1002	0.1024	0.0904	0.0903	0.0602	0.0923	0.0657	0.0548

5.2. *Bootstrap procedure.* In nonparametric estimation, asymptotic confidence intervals are not very accurate when the sample size is moderate. Bootstrap procedures are often used to bypass this problem. Asymptotic properties for bootstrapping homogeneous Markov chains can be found in Kulperger and Prakasa Rao [34] for the finite-state case and Athreya and Fuh [2] for the denumerable case. In particular, a natural idea is to generate replicates of the path of the Markov chain by using the estimation of the transition matrix. However in our case, the data are nonstationary and we combine this approach with the bootstrap scheme studied in Fryzlewicz et al. [26] for locally stationary ARCH processes. More precisely, for some $u \in (0, 1)$, an asymptotic confidence interval for Q_u can be obtained using the quantiles of the distribution of $\hat{Q}_u^* - \hat{Q}_u$, where

$$\hat{Q}_u^*(x, y) = \frac{\sum_{i=2}^n K_b(u - i/n) \mathbb{1}_{\{X_{i-1}^* = x, X_i^* = y\}}}{\sum_{i=2}^n K_b(u - j/n) \mathbb{1}_{\{X_{i-1}^* = x\}}}.$$

Here, $X_1^*, \ldots, X_n^*$ is, conditionally to the observations, a path of a Markov chain with transition matrix $\hat{Q}_u$. We set $\mathbb{P}^* = \mathbb{P}\left(\cdot | \sigma\left(X_{n,k} : 1 \leq k \leq n, k \geq 1\right)\right)$. Asymptotic validity of this bootstrap is justified by the following result. For simplicity, we only give a result for one entry of the stochastic matrix but a vectorial extension can be easily derived. The proof of the following result can be found in the supplementary material, Section 21.

PROPOSITION 5. *Suppose that Assumptions* **A1-A2** *hold true. Let K be a kernel supported on $[-1, 1]$, symmetric and of bounded variation. If $b \to 0$ and $nb^{1+\epsilon} \to \infty$ for some $\epsilon > 0$, then almost surely, the probability distribution of $\sqrt{nb}\left(\hat{Q}_u^*(x, y) - \hat{Q}_u(x, y)\right)$ under $\mathbb{P}^*$ converges to the Gaussian distribution of mean 0 and variance $\int_{-1}^1 K^2(v) dv \frac{Q_u(x,y)}{\pi_u(x)}\left(1 - Q_u(x, y)\right)$. Moreover if $nb^3 = O(1)$ and $u \mapsto Q_u$ is continuously differentiable, the distribution of $\sqrt{nb}\left(\hat{Q}_u^*(x, y) - \hat{Q}_u(x, y)\right)$ under $\mathbb{P}^*$ and $\sqrt{nb}\left(\hat{Q}_u(x, y) - Q_u(x, y)\right)$ under $\mathbb{P}$ are asymptotically equivalent.*

6. Application to the analysis of trading activity. Our real data illustration concerns the trading activity of six thinly traded shares at the
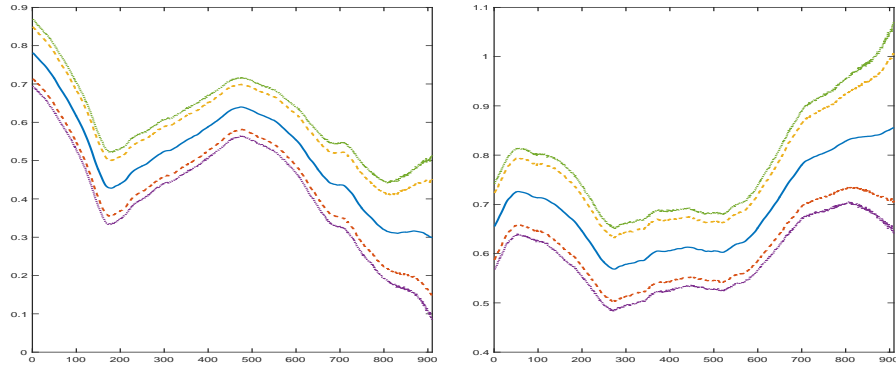
FIG 2. *Estimation of $u \mapsto Q_u(0,0)$ (left) and $u \mapsto Q_u(1,1)$ (right). The estimates are given by the full line and the dashed lines (dotted lines resp.) represent the bootstrap pointwise confidence intervals at level 80% (90% resp.) and which are estimated by using $B = 5000$ bootstrap samples of a stationary Markov chain with transition matrix $\hat{Q}_u$. The $x-$axis represents the time and the $y-$axis, the estimated local probability.*

Johannesburg Stock Exchange from 5th of October 1987 to 3rd of June 1991. These data are analyzed in Fokianos and Moysiadis [23] using stationary logistic and probit models. The data are binary, with a value equal to 1 if a trade has been recorded at time t and 0 otherwise. In Figure 1 given in Section 1, the function $u \mapsto \hat{\pi}_u(1)$ is represented for two shares which seems particularly inhomogeneous. While the probability to have a trade for the share "Anamint" follows a strong increase at the end of the period, that of the share "Broadcares" has the opposite behavior. In Fokianos and Moysiadis [23], the autocorrelograms of these two time series seems to exhibit significant correlations for large lags. For financial data, this kind of persistence is quite usual and often due to nonstationarity problems. See for instance Granger and Stărică [27] for a discussion of this phenomenon. We fit a time-inhomogeneous binary Markov chain to model the dynamic of the share "Anamint". An estimation of the diagonal elements of the stochastic matrix is given in Figure 2. Our approach suggests that the dynamic is strongly inhomogeneous.

One can also check that the graphs of the estimated local invariant probability in Figure 1 are compatible with the graph given in Fokianos and Moysiadis [23] with vertical bars for the presence of trading. The main advantage of time-inhomogeneous Markov chains is to get a statistical model which at the time is able to identify some trading patterns. Of course, one may think of using such models for prediction but this requires investigating higher order time-inhomogeneous Markov chains and probably parsimonious

versions of such models with for instance the locally stationary versions of probit/logit models. This is outside the scope of this paper.

7. Discussion. In this paper, we discussed various approaches for considering locally stationary versions of Markov chains models. The notion of local stationarity introduced in the literature offers a nice approach to deal with time-inhomogeneity but it is more adapted to continuous state space autoregressive Markov processes. Existing works exclude categorical data or integer-valued time series. We have defined a general notion of local stationarity based on a local approximation of the finite dimensional distributions using various probability metrics. This approach is quite flexible because various metrics can be used to define a locally stationary model. We now provide a guideline to precise what type of metric can be used to define a locally stationary version of a Markov chain model.

1. If the stationary version satisfies the Doeblin condition, then the total variation distance discussed in Section 2 is appropriate.
2. When the Doeblin condition is not satisfied but a drift and a small set condition can be obtained for the homogeneous Markov chain, the $V-$norm discussed in Section 4 could be used. Note that both approaches require the Markov kernels to be absolutely continuous with respect to a measure not depending on the parameters.
3. When a small set condition is too restrictive or cannot be checked, then a Wasserstein metric can be interesting when the state space is a Banach space. In this case, there is often a natural coupling of $(\delta_x Q, \delta_y Q)$ for the Markov kernel Q, as for autoregressive processes, to check the assumptions of Section 3.

Note that we did not develop statistical inference for the last case which generalizes a setup largely exploited in the literature. Let us also mention that the $\phi, \beta-$mixing coefficients are those already used for the stationary version of these Markov chains and the $\tau-$mixing coefficients were already discussed in Dedecker and Prieur [16] for some time-homogeneous Markov chains satisfying similar assumptions.

For the perspectives, bandwidth selection for the local contrast estimates studied in this paper, in the spirit of the recent work of Richter and Dahlhaus [39], could be investigated. Estimation of Markov-switching or SETAR models remain to do and are quite challenging. Another issue could be to investigate some processes involving a latent process defined by recursive equations, such as the Poisson GARCH and its variants, see for instance in Davis et al. [15], or the categorical time series studied in Moysiadis and Fokianos [37]. In this case, the small set condition is not satisfied and the contraction in

Wassertein metric is difficult to obtain or too restrictive.

References.

[1] M. Al Osh and A. Alzaid. Firs-order integer-valued autoregressive process. *J. Time Series Anal.*, 8:261–275, 1987.

[2] K.B. Athreya and C.D. Fuh. Bootstrapping markov chains: countable case. *J. Stat. Plan. Infer.*, 33:311–331, 1992.

[3] P.J. Avery and D.A. Henderson. Fitting markov chain models to discrete state series such as dna sequences. *J. R. Stat. Soc. Ser. C*, 48:53–61, 1999.

[4] N. Bartoli and P. Del Moral. *Simulation et algorithmes stochastiques: une introduction avec applications.* Editions Cépadù, 2001.

[5] P. Billingsley. *Statistical Inference for Markov Processes.* University of Chicago Press, Chicago, 1961.

[6] S. Birr, S. Volgushev, T. Kley, H. Dette, and M. Hallin. Quantile spectral analysis for locally stationary time series. *J. R. Stat. Soc. Ser. B*, 79(5):1619–1643, 2017.

[7] G. Bolch, S. Greiner, H. de Meer, and K.S. Trivedi. *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, 2nd Edition.* Wiley, 2006.

[8] R.C. Bradley. *Introduction to Strong Mixing Conditions, Vol 1.* Technical Report, Department of Mathematics, I.U. Bloomington, 2002.

[9] Richard C Bradley et al. Basic properties of strong mixing conditions. a survey and some open questions. *Probability surveys*, 2:107–144, 2005.

[10] D.R. Brillinger, P.A Morettin, R.A. Irizarry, and C. Chiann. Some wavelet-based analyses of markov chain data. *Signal Processing*, 80:1607–1627, 2000.

[11] Xi-Ren Cao. The maclaurin series for performance functions of markov chains. *Advances in Applied Probability*, 30(3):676–692, 1998.

[12] R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25:1–37, 1997.

[13] R. Dahlhaus and S. Subba Rao. Statistical inference for time-varying arch processes. *Ann. Statist.*, 34:1075–1114, 2006.

[14] R. Dahlhaus, S. Richter, and Wu W.B. Towards a general theory for non-linear locally stationary processes. *arXiv:1704.02860*, 2017.

[15] Richard A Davis, Scott H Holan, Robert Lund, and Nalini Ravishanker. *Handbook of discrete-valued time series.* CRC Press, 2016.

[16] J. Dedecker and C. Prieur. Coupling for $\tau-$dependent sequences and applications. *Journal of Theoretical Probability*, 17:861–885, 2004.

[17] R.L. Dobrushin. Central limit theorems for nonstationary markov chains. *Th. Prob. Appl.*, 1:329–383, 1956.

[18] R.L. Dobrushin. Prescribing a system of random variables by conditional distributions. *Th. Prob. Appl.*, 15:458–486, 1970.

[19] R. Douc, E. Moulines, and J.S. Rosenthal. Quantitative bounds on convergence of time-inhomogeneous markov chains. *Ann. Appl. Probab.*, 14(4):1643–1665, 2004.

[20] R. Douc, E. Moulines, and D. Stoffer. *Nonlinear Time Series.* Chapman and Hall, 2014.

[21] P. Doukhan. *Mixing. Properties and Examples.* Springer-Verlag, 1994.

[22] D. Ferré, L. Herv, and J. Ledoux. Regular perturbation of $v-$geometrically ergodic markov chains. *J. Appl. Prob.*, 50:184–194, 2013.

[23] K. Fokianos and T. Moysiadis. Binary time series models driven by a latent process. *Econometrics and Statistics*, 000:1–14, 2017.

[24] K. Fokianos, A. Rahbek, and D. Tjostheim. Poisson autoregression. *J. Amer. Statist. Assoc.*, 104:1430–1439, 2009.

[25] P. Fryzlewicz and S. Subba Rao. Mixing properties of arch and time-varying arch processes. *Bernoulli*, 17(1):320–346, 2011.

[26] P. Fryzlewicz, T. Sapatinas, and S. Subba Rao. Normalized least-squares estimation in time-varying arch models. *Ann. Statist.*, 36:742–786, 2008.

[27] C. Granger and C. Stărică. Nonstationarities in stock returns. *Review of Economics and Statistics*, 87:503–522, 2005.

[28] M. Hairer and J.C. Mattingly. Yet another look at harris' ergodic theorem for markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications IV*, volume 63, pages 109–117. Birkhuser,/Springer Basel AG, Basel, 2011.

[29] P. Hall and E. Bura. Nonparametric methods of inference for finite-state, inhomogeneous markov processes. *Bernoulli*, 10:919–938, 2004.

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

[31] Lajos Horváth and Piotr Kokoszka. *Inference for functional data with applications*, volume 200. Springer Science & Business Media, 2012.

[32] I.A. Ibragimov and Y.V. Linnik. *Independent and stationary sequences of random variables.* Wolters-Noordhoff, 1971.

[33] D. Jin-Guan and L. Yuan. The integer-valued autoregressive (inar(p)) model. *J. Time Series Anal.*, 12:129–142, 1991.

[34] R.J. Kulperger and B.L.S. Prakasa Rao. Bootstrapping a finite state markov chain. *Sankhyā*, 51:178–191, 1989.

[35] Guodong Li, Bo Guan, Wai Keung Li, and Philip L. H. Yu. Hysteretic autoregressive time series models. *Biometrika*, 102(3):717–723, 2015.

[36] S. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability 2nd.* Cambridge University Press New York, 2009.

[37] T. Moysiadis and K. Fokianos. On binary and categorical time series models with feedback. *Journal of Multivariate Analysis*, 131:209–228, 2014.

[38] B. Rajagopalan, U. Lall, and D.G. Tarboton. Nonhomogeneous markov model for daily precipitation. *Journal of Hydrologic Engineering*, 1:33–40, 1996.

[39] S. Richter and R. Dahlhaus. Cross validation for locally stationary processes. *arXiv:1705.10046*, 2017.

[40] L. Saloff-Coste and J. Zúñiga. Convergence of some time-inhomogeneous markov chains via spectral techniques. *Stochastic Process. Appl.*, 117:961–979, 2007.

[41] L. Saloff-Coste and J. Zúñiga. Merging for inhomogeneous finite markov chains, part ii: Nash and log-sobolev inequalities. *Ann. Probab.*, 39:1161–1203, 2011.

[42] R. R. Sarukkai. Link prediction and path analysis using markov chains. *Computer Networks*, 33:377–386, 2000.

[43] E. Seneta. *Non-negative Matrices and Markov Chains. Revised Printing.* Springer, 2006.

[44] J-P. Stockis, J. Franke, and J.T. Kamgaing. On geometric ergodicity of charme models. *Journal of Time Series Analysis*, 31(3):141–152, 2010.

[45] S. Subba Rao. On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in App. Probab.*, 38:1155–1172, 2006.

[46] Howell Tong. *Non-linear time series: a dynamical system approach.* Oxford University Press, 1990.

[47] L. Truquet. A perturbation analysis of some markov chains models with time-varying parameters. *arXiv :1610.09272*, 2017.

[48] N. Vergne. Drifting markov models with polynomial drift and applications to dna

sequences. *Stat. Appl. Genet. Molec. Biol.*, 7:1–45, 2008.

[49] C. Villani. *Optimal Transport. Old and New.* Springer, 2009.

[50] M. Vogt. Nonparametric regression for locally stationary time series. *Ann. Statist.*, 40:2601–2633, 2012.

[51] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods.* Springer, 1995.

[52] T. Zhang and W.B. Wu. Time-varying nonlinear regression models: nonparametric estimation and model selection. *Ann. Statist.*, 43:741–768, 2015.

ENSAI - Campus de Ker-Lann
Rue Blaise Pascal - BP 37203
35712 BRUZ cedex
FRANCE
E-mail: lionel.truquet@ensai.fr