Supplementary material for the paper: Local stationarity and time-inhomogeneous Markov chains

Lionel Truquet *†

In this notes, we provide the proofs of the results given in the paper as well as additional examples of locally stationary Markov chains. For the proofs, the number of the section is already given in the paper.

1 Proof of Theorem 1

We remind the reader that for a Markov kernel R on $(E, \mathcal{B}(E))$ and $\mu, \nu \in \mathcal{P}(E)$, $\|\mu R - \nu R\|_{TV} \leq c(R) \cdot \|\mu - \nu\|_{TV}$, where $c(R) = \sup_{(x,y)\in E} \|\delta_x R - \delta_y R\|_{TV} \in [0, 1]$. Then, under our assumptions, the application $T : \mathcal{P}(E) \to \mathcal{P}(E)$ defined by $T(\mu) = \mu Q_u^m$ is contracting and the existence and uniqueness of an invariant probability π_u easily follows from the fixed point theorem in a complete metric space.

We next check the first condition of Definition 1. The result is shown by induction. For $j = 1$, we have from assumption **A1**,

$$
\begin{aligned}
\|\pi_u - \pi_v\|_{TV} &\leq \|\pi_u Q_u^m - \pi_v Q_u^m\|_{TV} + \|\pi_v Q_u^m - \pi_v Q_v^m\|_{TV} \\
&\leq r\|\pi_u - \pi_v\|_{TV} + \sup_{x\in E} \|\delta_x Q_u^m - \delta_x Q_v^m\|_{TV}.
\end{aligned}
$$

Since for two Markov kernels R and $\widetilde{R}$ and $\mu, \nu \in \mathcal{P}(E)$, we have

$$
\|\mu R - \nu \widetilde{R}\|_{TV} \leq \sup_{x\in E} \|\delta_x R - \delta_x \widetilde{R}\|_{TV} + c\left(\widetilde{R}\right) \|\mu - \nu\|_{TV},
$$

we deduce from assumption **A2** that $\sup_{x\in E} \|\delta_x Q_u^m - \delta_x Q_v^m\|_{TV} \leq mL|u - v|$. This leads to the inequality $\|\pi_u - \pi_v\|_{TV} \leq mL(1-r)^{-1}|u-v|$ which gives the result for $j = 1$. If the continuity condition holds true for $j-1$, we note that $\pi_{u,j}(dx_1, \ldots, dx_{j-1}) = \pi_{u,j-1}(dx_1, \ldots, dx_{j-1}) Q_u(x_{j-1}, dx_j)$ and $\|\pi_{u,j} - \pi_{v,j}\|_{TV} \leq \sup_{x\in E} \|\delta_x Q_u - \delta_x Q_v\|_{TV} + \|\pi_{u,j-1} - \pi_{v,j-1}\|_{TV}$, which leads to the continuity of $u \mapsto \pi_{u,j}$. This justifies Condition 1 of Definition 1. Finally we prove the bound announced for $\|\pi_{k,j}^{(n)} - \pi_{u,j}\|_{TV}$. Note that this bound automatically implies Condition 2 of Definition 1. For $n \geq k \geq m$, we set $R_{k,m} = Q_{\frac{k-m+1}{n}} Q_{\frac{k-m+2}{n}} \cdots Q_{\frac{k}{n}}$. From Assumption **A2**, we have

$$
\sup_{x\in E} \|\delta_x R_{k,m} - \delta_x Q_u^m\|_{TV} \leq L \sum_{s=k-m+1}^{k} \left| u - \frac{s}{n} \right|.
$$

*UMR 6625 CNRS IRMAR, University of Rennes 1, Campus de Beaulieu, F-35042 Rennes Cedex, France and
†Campus de Ker-Lann, rue Blaise Pascal, BP 37203, 35172 Bruz cedex, France. *Email: lionel.truquet@ensai.fr.*

Now for $j=1$, we have

$$\begin{aligned}\|\pi_k^{(n)}-\pi_u\|_{TV} &\leq \|\pi_{k-m}^{(n)}R_{k,m}-\pi_{k-m}^{(n)}Q_u^m\|_{TV}+\|\pi_{k-m}^{(n)}Q_u^m-\pi_uQ_u^m\|_{TV}\\ &\leq L\sum_{s=k-m+1}^{k}\left|u-\frac{s}{n}\right|+r\|\pi_{k-m}^{(n)}-\pi_u\|_{TV}.\end{aligned}$$

Using the fact that is $s\leq 0$, $|u-s/n|\geq |u|$, we deduce that

$$\begin{aligned}\|\pi_k^{(n)}-\pi_u\|_{TV} &\leq L\sum_{\ell=0}^{\infty}r^{\ell}\sum_{s=k-(\ell+1)m+1}^{k-\ell m}\left|u-\frac{s}{n}\right|\\ &\leq Lm\sum_{\ell=0}^{\infty}\ell r^{\ell}\left|u-\frac{k}{n}\right|+\frac{1}{n}L\sum_{\ell=0}^{\infty}r^{\ell}\sum_{h=\ell m}^{(\ell+1)m-1}h.\end{aligned}$$

which gives the result for $j=1$. Next, using the same argument as for the continuity of the finite-dimensional distributions, we have

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

Hence the result easily follows by induction.□

2 An additional result for Section 2.3

In this section, we give a useful result for controlling the variance in the nonparametric kernel estimation of some expectations of finite-state Markov chains. We remind that $K:\mathbb{R}\rightarrow\mathbb{R}_+$ denotes a probability density, supported on $[-1,1]$ and of bounded variation. For $b=b_n\in(0,1)$, we set

$$e_i(u)=\frac{\frac{1}{nb}K\left(\frac{u-\frac{i}{n}}{b}\right)}{\frac{1}{nb}\sum_{j=\ell}^{n}K\left(\frac{u-\frac{j}{n}}{b}\right)},\quad u\in[0,1],\quad \ell\leq i\leq n$$

and

$$\hat{h}_u=\sum_{i=\ell}^{n}e_i(u)f\left(X_{n,i-\ell+1},\ldots,X_{n,i}\right).$$

The next proposition gives a uniform control of the variance part $\hat{h}_u-\mathbb{E}\hat{h}_u$.

Proposition 6. *Suppose that Assumptions* **A1-A2** *hold. Then, if $b\rightarrow 0$ and $nb^{1+\epsilon}\rightarrow\infty$ for some $\epsilon>0$,*

$$\sup_{u\in[0,1]}\left|\hat{h}_u-\mathbb{E}\hat{h}_u\right|=O_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{\sqrt{nb}}\right).$$

Proof of Proposition 6. We set $Y_{n,i} = f\left(X_{n,i-\ell+1}, \ldots, X_{n,i}\right)$. First, note that the triangular array $\left(Y_{n,i}\right)_{1 \leq i \leq n}$ is $\phi-$mixing (and then $\alpha-$mixing) with $\phi_n(j) \leq \widetilde{C}\rho^{j-\ell}$ where $\widetilde{C}$ is a positive constant. We have $[0,1] = \cup_{s=1}^{k+1} I_s$ where k is the integer part of $1/b$, $I_s = ((s-1)b, sb]$ for $1 \leq s \leq k$ and $I_{k+1} = (kb, 1]$. We set $S_0^{(n)} = 0$ and if $\ell \leq i \leq n$, $S_i^{(n)} = \sum_{s=\ell}^{i} Z_s^{(n)}$, where $Z_s^{(n)} = Y_{n,s} - \mathbb{E}Y_{n,s}$. Then for $\ell \leq j \leq j+k \leq n$, we have

$$\begin{aligned}\left|\sum_{i=j}^{j+k} e_i(u) Z_i^{(n)}\right| &\leq e_j(u) \cdot \left|S_{j-1}^{(n)}\right| + e_{j+k}(u) \cdot \left|S_{j+k}^{(n)}\right| + \sum_{i=j}^{j+k-1} \left|e_i(u) - e_{i-1}(u)\right| \cdot \left|S_i^{(n)}\right| \\ &\leq \frac{C''}{nb} \max_{j-1 \leq i \leq j+k} \left|S_i^{(n)}\right|.\end{aligned}$$

This gives the bound

$$\begin{aligned}\max_{u \in [0,1]} \left|\sum_{i=1}^{n} e_i(u) Z_i^{(n)}\right| &\leq \max_{1 \leq s \leq k} \max_{u \in I_s} \left|\sum_{n(s-2)b \leq i \leq n(s+1)b} e_i(u) Z_i^{(n)}\right| \\ &\leq \frac{C''}{nb} \max_{2 \leq s \leq k+1} \max_{n(s-2)b-1 \leq i \leq n(s+1)b} \left|S_i^{(n)}\right|.\end{aligned}$$

We will use the exponential inequality for strong mixing sequences given in Rio (1999), Theorem 6.1 (see also Rio (2013), Theorem 6.1). This inequality guarantees that for any integer q, we have

$$\mathbb{P}\left(\max_{n(s-2)b-1 \leq i \leq n(s+1)b} \left|S_i^{(n)}\right| \geq F\lambda\right) \leq G \exp\left(-\frac{\lambda}{2q\|f\|_\infty} \log\left(1 + K\frac{\lambda q}{nb}\right)\right) + Mnb\frac{\rho^q}{\lambda}, \tag{1}$$

where F, G, K, M are three positive real numbers not depending on n and s and $\lambda \geq q\|f\|_\infty$. We have $k = O\left(b^{-1}\right)$ and setting $q \approx \frac{\sqrt{nb}}{\sqrt{\log n}}$ and $\lambda = \lambda'\sqrt{nb\log n}$, we have for λ' large enough

$$\mathbb{P}\left(\max_{u \in [0,1]} \left|\sum_{i=1}^{n} e_i(u) Z_i^{(n)}\right| > \frac{F\lambda}{nb}\right) = O_{\mathbb{P}}\left(\frac{1}{bn^{\frac{1}{1+\epsilon}}} + \frac{\sqrt{nb}}{b\sqrt{\log(n)}}\rho^{\frac{\sqrt{nb}}{\log(n)}}\right).$$

Then the result follows from the bandwidth conditions.$\square$

3 Proof of the results of Section 2

3.1 Proof of Proposition 1

Using the Markov property, we have

$$\phi_n(j) \leq \max_{1 \leq i \leq n-j} \sup_{0 \leq f \leq 1} \left\|\mathbb{E}\left(f\left(X_{n,i+j}\right) | X_{n,i}\right) - \mathbb{E}\left(f\left(X_{n,i}\right)\right)\right\|_\infty,$$

where for a random variable Y, $\|Y\|_\infty$ denotes its infinite norm. We first consider $\epsilon > 0$ such that $\alpha = 2mL\epsilon + r < 1$. Assume first that $n \geq \frac{m}{\epsilon}$. In the proof of Theorem 1, we have shown that if for $m \leq k \leq n$ and $R_{k,m} = Q_{\frac{k-m+1}{n}} \cdots Q_{\frac{k}{n}}$,

$$\sup_{x \in E} \|\delta_x R_{k,m} - \delta_x Q_{\frac{k}{n}}^m\|_{TV} \leq L \sum_{s=k-m+1}^{k} \left|u - \frac{s}{n}\right| \leq mL\epsilon.$$

Then, from Assumption **A1** and the triangular inequality, we get

$$\sup_{x,y\in E}\|\delta_x R_{k,m}-\delta_y R_{k,m}\|_{TV}\leq \alpha.$$

Now if $j=tm+s$ for two positive integers t,s, we get

$$\|\delta_{X_{n,k-j}}Q_{\frac{k-j+1}{n}}\cdots Q_{\frac{k}{n}}-\pi^{(n)}_{k-j}Q_{\frac{k-j+1}{n}}\cdots Q_{\frac{k}{n}}\|_{TV}\leq \alpha^t.$$

We easily deduce the bound on ϕ_n by taking the infinite norm for the left-hand term of the previous inequality. Now, if $n<\frac{m}{\epsilon}$, one can use the bound $\phi_n(j)\leq 1$. Setting $\rho=\alpha^{1/m}$, this leads to the result with an appropriate choice of C, e.g $C=\max\left\{\rho^{1-\frac{1}{\epsilon}},\alpha^{-1}\right\}$.□

3.2 Proof of Corollary 1

From the inequality

$$\|\delta_x Q_u^m-\delta_y Q_u^m\|_{TV}=1-\sum_{z\in E}Q_u^m(x,z)\wedge Q_u^m(y,z)\leq 1-|E|\cdot\inf_{x,y\in E}Q_u^m(x,y),$$

assumption **A1** is satisfied as soon as $\inf_{u\in[0,1],(x,y)\in E^2}Q_u^m(x,y)>0$. From aperiodicity and irreducibility, it is well known that for each $u\in[0,1]$,

$$m_u=\inf\left\{k\geq 1:\min_{(x,y)\in E^2}Q_u^k(x,y)>0\right\}<\infty.$$

By continuity of the application $u\mapsto Q_u$, the sets $\mathcal{O}_u=\{v\in[0,1]:Q_v^{m_u}>0\}$ are open subsets of $[0,1]$. Using a compactness argument, the interval $[0,1]$ can be covered by finitely many $\mathcal{O}_u$, say $\mathcal{O}_{u_1},\ldots,\mathcal{O}_{u_d}$. Then Assumption **A1** is satisfied with $m=\max_{1\leq i\leq d}m_{u_i}$. Assumption **A2** is automatically satisfied. Then Theorem 1 and Proposition 1 apply.□

3.3 Asymptotic properties for finite-state Markov chains

We will prove the following result.

Theorem 5. *Suppose that Assumptions* **A1-A2** *hold and that for a given $\epsilon>0$, $b\to 0$ and $nb^{1+\epsilon}\to\infty$.*

1. *For $(x,y)\in E^2$, we have*

$$\sup_{u\in[0,1]}\left[|\mathbb{E}\hat{\pi}_u(x)-\pi_u(x)|+\left|\frac{\mathbb{E}\hat{\pi}_{u,2}(x,y)}{\mathbb{E}\hat{\pi}_u(x)}-Q_u(x,y)\right|\right]=O(b)\tag{2}$$

and

$$\sup_{u\in[0,1]}|\hat{\pi}_u(x)-\mathbb{E}\hat{\pi}_u(x)|=O\left(\frac{\sqrt{\log(n)}}{\sqrt{nb}}\right),\tag{3}$$

$$\sup_{u\in[0,1]}\left|\hat{Q}_u(x,y)-\frac{\mathbb{E}\hat{\pi}_{u,2}(x,y)}{\mathbb{E}\hat{\pi}_u(x)}\right|=O\left(\frac{\sqrt{\log(n)}}{\sqrt{nb}}\right).$$

2. *For* $(u,x)\in(0,1)\times E$, *the vector* $\left(\sqrt{nb}\left[\hat{\pi}_u(x)-\mathbb{E}\hat{\pi}_u(x)\right]\right)_{x\in E}$ *is asymptotically Gaussian with mean* 0 *and covariance* $\Sigma_u^{(1)}:E\times E\to\mathbb{R}$ *defined by*

$$\Sigma_u^{(1)}=\int K^2(x)dx\cdot\left[\Gamma_u(0)+\sum_{j\geq 1}\left(\Gamma_u(j)+\Gamma_u(j)'\right)\right],$$

where $\Gamma_u(j)_{x,y}=\pi_u(x)Q_u^j(x,y)-\pi_u(x)\pi_u(y)$.

3. *For* $(u,x,y)\in(0,1)\times E^2$, *the vector*

$$\sqrt{nb}\left(\hat{Q}_u(x,y)-\frac{\mathbb{E}\hat{\pi}_{u,2}(x,y)}{\mathbb{E}\hat{\pi}_u(x)}\right)_{(x,y)\in E^2}$$

is asymptotically Gaussian with mean 0 *and covariance* $\Sigma^{(2)}:E^2\times E^2\to\mathbb{R}$ *defined by*

$$\Sigma_u^{(2)}\left((x,y),(x',y')\right)=\frac{\int K^2(x)dx\cdot Q_u(x,y)}{\pi_u(x)}\left[\mathbb{1}_{y=y'}-Q_u(x',y')\right]\mathbb{1}_{x=x'}.$$

Proof of Theorem 5

Proof of point 1. For the control of the bias, note that

$$\mathbb{E}\hat{\pi}_{u,2}(x,y)-\pi_{u,2}(x,y)=\sum_{i=1}^{n-1}e_i(u)\left[\pi_{i,2}^{(n)}(x,y)-\pi_{u,2}(x,y)\right].$$

Since $e_i(u)=0$ if $|u-i/n|>b$, Theorem 1 ensures that

$$\sup_{u\in[0,1]}\left|\mathbb{E}\hat{\pi}_{u,2}(x,y)-\pi_{u,2}(x,y)\right|=O\left(b+\frac{1}{n}\right)=O\left(b\right).$$

By summation on y, we deduce the first bound and using the fact that $\min_{u\in[0,1]}\pi_u(x)>0$, we deduce that $\max_{u\in[0,1]}\frac{1}{\mathbb{E}\hat{\pi}_u(x)}=O_{\mathbb{P}}(1)$ and the second bound follows.
For the variance terms, we use Proposition 6 which ensures the first bound as well as $\max_{u\in[0,1]}\frac{1}{\hat{\pi}_u(x)}=O_{\mathbb{P}}(1)$. This gives also the second bound.□

Proof of point 2. The proof is a simple consequence of Proposition 5 (2.) given in the paper. Indeed, from the proof of Corollary 1 of the paper, it is shown that there exists a positive integer m such that $\min_{x,y\in E}\min_{u\in[0,1]}Q_u^m(x,y)>0$. Using the uniform continuity of the application $u\mapsto Q_u$, one can check that all the assumptions of Theorem 4 are satisfied by choosing $\epsilon>0$ small enough, $V\equiv 1$ and $\widetilde{V}$ constant. Then setting $Z_{n,i}=\sum_{x\in E}\lambda_x\mathbb{1}_{\{X_{n,i}=x\}}$ for some real numbers $\lambda_x,x\in E$, the result follows directly from Proposition 7 and the Cramér-Wold device.□

Proof of point 3. Let

$$Z_n(x,y) = \frac{\sqrt{nb}}{\hat{\pi}_u(x)} \sum_{i=2}^n D_{n,i}(x,y)$$

where

$$D_{n,i}(x,y) = e_i(u) \left[\mathbb{1}_{\{X_{n,i-1}=x, X_{n,i}=y\}} - Q_{\frac{i}{n}}(x,y) \mathbb{1}_{\{X_{n,i-1}=x\}} \right]$$

is a martingale increment bounded by $(nb)^{-1}$ (up to a constant). We set $\delta_{n,i}(x,y) = 1$ if $X_{n,i-1} = x, X_{n,i} = y$ and 0 otherwise and $\delta_i^{(u)}(x,y)$ is defined in the same way but with the stationary approximation. Using the classical Lindeberg central limit theorem for martingales, the sum $\sqrt{nb}\sum_{i=1}^{n-1} \left[D_{n,i}(x,y)\right]_{x,y\in E}$ is asymptotically a Gaussian vector with mean 0 and variance matrix Σ defined by

$$\begin{aligned}
& \Sigma\left((x,y),(x',y')\right) \\
= & \lim_{n\to\infty} nb \sum_{i=2}^n e_i(u)^2 \mathrm{Cov}\left[\delta_{n,i}(x,y) - Q_{\frac{i}{n}}(x,y)\mathbb{1}_{\{X_{n,i-1}=x\}}, \delta_{n,i}(x',y') - Q_{\frac{i}{n}}(x',y')\mathbb{1}_{\{X_{n,i-1}=x'\}}\right] \\
= & \lim_{n\to\infty} nb \sum_{i=2}^n e_i(u)^2 \mathrm{Cov}\left[\delta_i^{(u)}(x,y) - Q_u(x,y)\mathbb{1}_{\{X_{i-1}(u)=x\}}, \delta_i^{(u)}(x',y') - Q_u(x',y')\mathbb{1}_{\{X_{i-1}(u)=x'\}}\right] \\
= & \int K^2(z)dz \cdot \mathbb{P}\left(X_1(u)=x, X_2(u)=y\right) \cdot \left[\mathbb{1}_{y=y'} - Q_u(x',y')\right] \mathbb{1}_{x=x'}.
\end{aligned}$$

In the previous equalities, we have used Theorem 1, the continuity properties of the transition matrix and the limits

$$\lim_{n\to\infty} \frac{1}{nb}\sum_{i=1}^{n-1} K\left(\frac{u-i/n}{b}\right) = \int K(z)dz = 1, \quad \lim_{n\to\infty} \frac{1}{nb}\sum_{i=1}^{n-1} K^2\left(\frac{u-i/n}{b}\right) = \int K(z)^2 dz.$$

We deduce that the vector $\left[Z_n(x,y)\right]_{x,y\in E}$ is asymptotically Gaussian with mean zero and covariance matrix $\Sigma_u^{(2)}$.

Then it remains to show that for each $(x,y)\in E^2$,

$$\sqrt{nb}\left[\sum_{i=2}^n e_i(u)\frac{\mathbb{1}_{\{X_{n,i-1}=x\}}Q_{\frac{i}{n}}(x,y)}{\hat{\pi}_u(x)} - \frac{\mathbb{E}\hat{\pi}_{u,2}(x,y)}{\mathbb{E}\hat{\pi}_u(x)}\right] = o_{\mathbb{P}}(1). \tag{4}$$

To show (4), we use the decomposition

$$\begin{aligned}
& \sqrt{nb}\left[\sum_{i=2}^n e_i(u)\frac{\mathbb{1}_{\{X_{n,i-1}=x\}}Q_{\frac{i}{n}}(x,y)}{\hat{\pi}_u(x)} - \frac{\mathbb{E}\hat{\pi}_{u,2}(x,y)}{\mathbb{E}\hat{\pi}_u(x)}\right] \\
= & \frac{\sqrt{nb}}{\hat{\pi}_u(x)}\sum_{i=2}^n e_i(u)\left(\mathbb{1}_{\{X_{n,i-1}=x\}} - \pi_{i-1}^{(n)}(x)\right)\cdot\left(Q_{\frac{i}{n}}(x,y) - Q_u(x,y)\right) \\
+ & \sqrt{nb}\sum_{i=2}^n e_i(u)\pi_{i-1}^{(n)}(x)\left(Q_{\frac{i}{n}}(x,y) - Q_u(x,y)\right)\frac{\pi_u(x) - \hat{\pi}_u(x)}{\hat{\pi}_u(x)\pi_u(x)} \\
= & \frac{A_n}{\hat{\pi}_u(x)} + B_n\frac{\pi_u(x) - \hat{\pi}_u(x)}{\hat{\pi}_u(x)\pi_u(x)}.
\end{aligned}$$

Since the kernel K has a compact support and $u \mapsto Q_u(x,y)$ is Lipschitz continuous, we have $B_n = O\left(\sqrt{nb}b\right)$. Moreover, using covariance inequalities, we have $\mathrm{Var}\left(A_n\right) = O\left(b^2\right)$. Then (4) follows from $\hat{\pi}_u(x) - \pi_u(x) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nb}}\right)$ and $\frac{1}{\hat{\pi}_u(x)} = O_{\mathbb{P}}(1)$. The proof of point 3 is now complete. □

4 Geometric ergodicity result for Section 3

Proposition 7. *Assume that assumptions* **B1-B3** *hold true.*

1. *For all $u \in [0,1]$, the Markov chain of transition Q_u has a unique invariant probability distribution denoted by π_u. Moreover for all initial probability distribution $\mu \in \mathcal{P}_p(E)$, we have for $n = mj + s$*

$$W_p\left(\mu Q_u^n, \pi_u\right) \leq C_1^s r^j \left[\left(\int d(x,x_0)^p \mu(dx)\right)^{1/p} + \kappa_2\right],$$

where $\kappa_2 = \sup_{u\in[0,1]}\left(\int d(x,x_0)^p \pi_u(dx)\right)^{1/p} < \infty$.

2. *If $u, v \in [0,1]$, we have*

$$W_p(\pi_u, \pi_v) \leq \frac{C_2|u-v|}{1-r}\left[mC_1^{m-1}\kappa_2 + \sum_{j=0}^{m-1} C_1^j \kappa_1(m-j-1)\right],$$

where $\kappa_1(j) = \sup_{u\in[0,1]}\left(\int d(x,x_0)^p Q_u^j(x_0,dx)\right)^{1/p} < \infty$.

Proof of Proposition 7 We first show that the quantities $\kappa_1(j)$ are finite. We set $q_j = \left(\int (1+d(x,x_0))^p\, Q_0^j(x_0,dx)\right)^{1/p}$. If $j \geq 1$, we have, using Lemma 4 (2.),

$$\begin{aligned}
& W_p\left(\delta_{x_0} Q_u^j, \delta_{x_0} Q_0^j\right) \\
\leq\; & W_p\left(\delta_{x_0} Q_u^j, \delta_{x_0} Q_0^{j-1} Q_u\right) + W_p\left(\delta_{x_0} Q_0^{j-1} Q_u, \delta_{x_0} Q_0^j\right) \\
=\; & C_1 W_p\left(\delta_{x_0} Q_u^{j-1}, \delta_{x_0} Q_0^{j-1}\right) + C_2 |u|^\kappa q_{j-1}.
\end{aligned}$$

We obtain

$$W_p\left(\delta_{x_0} Q_u^j, \delta_{x_0} Q_0^j\right) \leq C_2 \sum_{s=0}^{j-1} C_1^s q_{j-s-1}. \tag{5}$$

Using Lemma 3 with the function $f(x) = 1 + d(x,x_0)$, we get

$$\kappa_1(j) \leq q_j + C_2 \sum_{s=0}^{j-1} C_1^s q_{j-s-1}.$$

1. The existence and uniqueness of an invariant probability $\pi_u \in \mathcal{P}_p$ easily follows from the fixed point theorem for a contracting application in the complete metric space $(\mathcal{P}_p, W_p)$.
   Before proving the geometric convergence, let us show that the quantity κ_2 is finite. We have, using Lemma 4,

$$\begin{aligned} W_p(\pi_u, \pi_0) &\leq W_p\left(\pi_u Q_u^m, \pi_0 Q_u^m\right) + W_p\left(\pi_0 Q_u^m, \pi_0 Q_0^m\right) \\ &\leq r W_p\left(\pi_u, \pi_0\right) + \left(\int W_p^p\left(\delta_x Q_u^m, \delta_x Q_0^m\right) \pi_0(dx)\right)^{1/p}. \end{aligned}$$

   Using (5) and Lemma 4, we have

$$\begin{aligned} W_p\left(\delta_x Q_u^m, \delta_x Q_0^m\right) &\leq W_p\left(\delta_x Q_u^m, \delta_{x_0} Q_u^m\right) + W_p\left(\delta_{x_0} Q_u^m, \delta_{x_0} Q_0^m\right) + W_p\left(\delta_x Q_0^m, \delta_{x_0} Q_0^m\right) \\ &\leq 2r d(x, x_0) + C_2 \sum_{s=0}^{m-1} C_1^j q_{m-s-1}. \end{aligned}$$

   From the previous bound, we easily deduce the existence of a real number $D > 0$, not depending on u, such that $W_p(\pi_u, \pi_0) \leq \frac{D}{1-r}$. Using Lemma 3, we get

$$\kappa_2 \leq \frac{D}{1-r} + \left(\int d(x, x_0)^p \pi_0(dx)\right)^{1/p},$$

   which is finite.
   Now, the geometric convergence is a consequence of the inequality

$$W_p\left(\mu Q_u^n, \pi_u Q_u^n\right) \leq C_1^s r^j W_p(\mu, \pi_u) \leq C_1^s r^j \left[\left(\int d(x, x_0)^p \mu(dx)\right)^{1/p} + \kappa_2\right].$$

   Finally, let ν be an invariant probability for P_u (not necessarily in $\mathcal{P}_p$). Let $f : E \to \mathbb{R}$ be an element of $\mathcal{C}_b(E)$ (the set of real-valued, continuous and bounded functions defined on E). Since convergence in Wasserstein metric implies weak convergence, we have from the geometric ergodicity $\lim_{n \to \infty} Q_u^n f(x) = \pi_u f$ for all $x \in E$. Hence, using the Lebesgue theorem, we have

$$\nu f = \nu Q_u^n f = \int \nu(dx) Q_u^n f(x) \to \pi_u f$$

   which shows the unicity of the invariant measure.

2. Proceeding as for the previous point, we have

$$W_p(\pi_u, \pi_v) \leq r W_p(\pi_u, \pi_v) + \left(\int W_p^p\left(\delta_x Q_u^m, \delta_x Q_v^m\right) \pi_v(dx)\right)^{1/p}. \tag{6}$$

   However,

$$\begin{aligned} W_p\left(\delta_x Q_u^m, \delta_x Q_v^m\right) &\leq C_1 W_p\left(\delta_x Q_u^{m-1}, \delta_x Q_v^{m-1}\right) + C_2 |u - v| \left(\int \left[1 + d(y, x_0)\right]^p Q_v^{m-1}(x, dy)\right)^{1/p} \\ &\leq C_1 W_p\left(\delta_x Q_u^{m-1}, \delta_x Q_v^{m-1}\right) + C_2 |u - v| \left(\kappa_1(m-1) + C_1^{m-1} d(x, x_0)\right). \end{aligned}$$

We deduce that

$$W_p\left(\delta_x Q_u^m, \delta_x Q_v^m\right) \leq C_2|u-v|\left(\sum_{j=0}^{m-1} C_1^j \kappa_1(m-j-1) + mC_1^{m-1} d(x,x_0)\right).$$

Reporting the last bound in (6), we get the result.□

5 Proof of Theorem 2

1. We show the result by induction and first consider the case $j=1$. For $k \leq n$, let $Q_{k,m}$ be the probability kernel $Q_{\frac{k-m+1}{n}} \cdots Q_{\frac{k}{n}}$. We have

$$\begin{aligned} W_p\left(\pi_k^{(n)}, \pi_u\right) &= W_p\left(\pi_{k-m}^{(n)} Q_{k,m}, \pi_u Q_u^m\right) \\ &\leq W_p\left(\pi_{k-m}^{(n)} Q_{k,m}, \pi_{k-m}^{(n)} Q_u^m\right) + W_p\left(\pi_{k-m}^{(n)} Q_u^m, \pi_u Q_u^m\right) \\ &\leq rW_p\left(\pi_{k-m}^{(n)}, \pi_u\right) + \left(\int W_p^p\left(\delta_x Q_{k,m}, \delta_x Q_u^m\right) \pi_{k-m}^{(n)}(dx)\right)^{1/p}. \end{aligned}$$

From Lemma 5, we have

$$W_p\left(\delta_x Q_{k,m}, \delta_x Q_u^m\right) \leq \sum_{s=0}^{m-1} C_1^s C_2 \left[\int (1+d(y,x_0))^p \delta_x Q_{\frac{k-m+1}{n}} \cdots Q_{\frac{k-s-1}{n}}(dy)\right]^{1/p} \left|u - \frac{k-s}{n}\right|.$$

First we note that from our assumptions and using Lemma 3 with the function $f(x) = 1 + d(x,x_0)$, we have

$$\left[\delta_x Q_u f^p\right]^{1/p} \leq \left[\delta_{x_0} Q_u f^p\right]^{1/p} + C_1 d(x,x_0) \leq \left(1 + \kappa_1(1) + C_1\right) f(x),$$

where κ_1 is defined in Proposition 2 in the paper. Then we get $\sup_{u\in[0,1]} \delta_x Q_u f^p \leq C_3^p f^p(x)$, where $C_3 = 1 + \kappa_1(1) + C_1$. This yields to the inequality

$$W_p\left(\delta_x Q_{k,m}, \delta_x Q_u^m\right) \leq \sum_{s=0}^{m-1} C_1^s C_2 C_3^{m-s-1} \left|u - \frac{k-s}{n}\right| f(x).$$

Then we obtain

$$W_p\left(\pi_k^{(n)}, \pi_u\right) \leq rW_p\left(\pi_{k-m}^{(n)}, \pi_u\right) + \sum_{s=0}^{m-1} C_1^s C_2 C_3^{m-s-1} \left|u - \frac{k-s}{n}\right| \left(\pi_{k-m}^{(n)} f^p\right)^{1/p}.$$

Then the result will easily follow if we prove that $\sup_{n,k\leq n} \pi_k^{(n)} f^p < \infty$. Setting $c_k = W_p\left(\pi_k^{(n)}, \pi_{\frac{k}{n}}\right)$ and $C_4 = \sum_{s=0}^{m-1}(s+1) C_1^{s+1} C_2 C_3^{m-s-1}$ and using our previous inequality, we have

$$\begin{aligned} c_k &\leq rW_p\left(\pi_{k-m}^{(n)}, \pi_{\frac{k}{n}}\right) + \frac{C_4}{n}\left(\pi_{k-m}^{(n)} f^p\right)^{1/p} \\ &\leq \left(r + \frac{C_4}{n}\right) c_{k-m} + rW_p\left(\pi_{\frac{k-m}{n}}, \pi_{\frac{k}{n}}\right) + \frac{C_4}{n}(1+\kappa_2). \end{aligned}$$

Then, if n_0 is such that for all $n \geq n_0$, $r+\frac{C_4}{n} < 1$, the last inequality, Proposition 2 and Lemma 3 guarantee that $\sup_{n\geq n_0, 1\leq k\leq n} \pi_k^{(n)} f^p$ is finite and only depends on $p, d, r, C_1, C_2, \kappa_1(1), \ldots, \kappa_1(m), \kappa_2$. Moreover if $n \leq n_0$, we have $\left(\pi_k^{(n)} f^p\right)^{1/p} \leq (C_4+1)^{n_0} \left(\pi_0 f^p\right)^{1/p}$. This concludes the proof for the case $j=1$.

2. Now for $j \geq 2$, we define a coupling of $\left(\pi_{k,j}^{(n)}, \pi_{u,j}\right)$ as follows. First we consider an optimal coupling $\Gamma_{u,j-1}^{(k,n)}$ of $\left(\pi_{k,j-1}^{(n)}, \pi_{u,j-1}\right)$, and for each $(x,y) \in E^2$, we define an optimal coupling $\Delta_{x,y,j,u}^{(k,n)}$ of $\left(\delta_x Q_{\frac{k+j}{n}}, \delta_y Q_u\right)$. From Villani (2009), Corollary 5.22, it is possible to choose this optimal coupling such that the application $(x,y) \mapsto \Delta_{x,y,j,u}^{(k,n)}$ is measurable. Now we define

$$\Gamma_{u,j}^{(k,n)}(dx_1, dy_1, \ldots, dx_j, dy_j) = \Delta_{x_{j-1},y_{j-1},j,u}^{(k,n)}(dx_j, dy_j)\Gamma_{u,j-1}^{(k,n)}(dx_1, dy_1, \ldots, dx_{j-1}, dy_{j-1}).$$

Then we easily deduce that

$$W_p^p\left(\pi_{k,j}^{(n)}, \pi_{u,j}\right) \leq W_p^p\left(\pi_{k,j-1}^{(n)}, \pi_{u,j-1}\right) + \int W_p^p\left(\delta_{x_{j-1}} Q_{\frac{k+j}{n}}, \delta_{y_{j-1}} Q_u\right) \Gamma_{u,j-1}^{(n,k)}\left(dx_1, dy_1, \ldots, dx_{j-1}, dy_{j-1}\right).$$

Since

$$W_p\left(\delta_{x_{j-1}} Q_{\frac{k+j}{n}}, \delta_{y_{j-1}} Q_u\right) \leq C_1 d(x_{j-1}, y_{j-1}) + C_2\left[1 + d(y_{j-1}, x_0)\right]\left|u - \frac{k+j}{n}\right|.$$

This leads to

$$W_p\left(\pi_{k,j}^{(n)}, \pi_{u,j}\right) \leq (1+C_1) W_p\left(\pi_{k,j-1}^{(n)}, \pi_{u,j-1}\right) + C_2(1+\kappa_2)\left|u - \frac{k+j}{n}\right|.$$

The results follows by a finite induction.

Finally, note that Condition 1 of Definition 1 of the paper follows from induction and the point 2 of Proposition 2, because using the same type of arguments, we have

$$W_p\left(\pi_{v,j}, \pi_{u,j}\right) \leq (1+C_1) W_p\left(\pi_{v,j-1}, \pi_{u,j-1}\right) + C_2(1+\kappa_2)\left|u - v\right|.$$

The proof of the Theorem is now complete.□

6 Local stationarity for the functional time series example

For a square integrable function f, we remind that $\mathbb{E}\left|\int_0^1 f(s) dB_k(s)\right|^2 = \int_0^1 f(s)^2 ds$. Here we set $x_0 = 0$. For a borelian set A and $x \in E$ we set

$$Q_u(x, A) = \mathbb{P}\left(\int a(\cdot, s) x(s) ds + \int \sigma(\cdot, s) dB_1(s) \in A\right).$$

Using the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, we have

$$\begin{aligned}
\int \|y\|^2 \delta_x Q_u(dy) &\leq 2\left[\int_0^1 \left|\int_0^1 a_u(t,s) x(s) ds\right|^2 dt + \int_0^1 \mathbb{E}\left|\int_0^1 \sigma_u(t,s) dB_1(s)\right|^2 dt\right] \\
&\leq 2\left[\int_0^1 \int_0^1 a_u^2(t,s) dt ds \|x\|^2 + \int_0^1 \int_0^1 \sigma_u(t,s) ds dt\right].
\end{aligned}$$

From the integrability assumptions, we get **B1**. Next, using Cauchy-Schwarz inequality,

$$
\begin{aligned}
W_2^2\left(\delta_x Q_u, \delta_y Q_u\right) &= \int_0^1 \left|\int_0^1 a_u(t,s)(x(s)-y(s))ds\right|^2 dt \\
&\leq \int_0^1 \int_0^1 a_u^2(t,s)dsdt \cdot \|x-y\|^2.
\end{aligned}
$$

Assumption **B2** follows from the assumption made on the kernel a. Finally, using similar arguments, we have

$$
W_2^2\left(\delta_x Q_u, \delta_x Q_v\right) \leq 2C^2|u-v|^2\left(1+\|x\|^2\right)
$$

and **B3** is also satisfied.

7 Mixing conditions for Markov chains contracting in the Wasserstein metric

For the Markov chains introduced in Section 3 of the paper, the $\tau-$mixing coefficients introduced and studied in Dedecker and Prieur (2004) are adpated to our triangular arrays. When this coefficient, which has been introduced for Banach spaces E, has a suitable decay, standard limit theorems can be obtained for the random sequence under study. The advantage of this mixing coefficient is to not require regularity conditions on the noise distributions of dynamical systems. We refer the reader to Dedecker and Prieur (2004) for many examples of random sequences for which this coefficient can be easily controlled. For our locally stationary Markov chains with contracting Markov kernels in Wasserstein metrics, the $\tau-$dependence is adapted and will replace the $\phi-$mixing coefficient of Section 2.

In the sequel, we denote by $\Lambda_1(E)$ the set of $1-$Lipschitz functions from E to $\mathbb{R}$. If $j \leq n-1$ and $1 \leq i \leq n-j$, we set

$$
U_{i,j}^{(n)} = \sup\left\{\left|\mathbb{E}\left[f\left(X_{n,i+j}|X_{n,i}\right)\right] - \mathbb{E}\left[f\left(X_{n,i+j}\right)\right]\right| : \quad f \in \Lambda_1(E)\right\},
$$

the τ_n-mixing coefficient for the sequence $\left(X_{n,k}\right)_{1\leq k\leq n}$ is defined by

$$
\tau_n(j) = \sup_{1\leq i\leq n-j} \mathbb{E}\left[U_{i,j}^{(n)}\right].
$$

Note that, if $\widetilde{X}_{n,i+j}$ denotes a copy of $X_{n,i+j}$, independent from $X_{n,i}$, we have the bound

$$
\mathbb{E}\left[U_{i,j}^{(n)}\right] \leq \mathbb{E}d\left(X_{n,i+j}, \widetilde{X}_{n,i+j}\right).
$$

This bound is particularly useful for bounding the coefficients τ_n. Let us also note that by the Kantorovitz-Rubinstein theorem (see Villani (2009), Remark 5.16), the random variable $U_{i,j}^{(n)}$ is the Wasserstein distance of order 1 between the probability distribution of $X_{n,i+j}$ and the conditional distribution $X_{n,i+1}|X_{n,i}$.

The following assumption, which strengthens the first part of Assumption **B2** in the case $m \geq 2$ and $p = 1$, will be needed. In the sequel, for a positive real number ϵ, we set

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

B4 There exists a positive real number ϵ such that for all $(u, u_1, \ldots, u_m) \in [0,1]^{m+1}$ satisfying $|u_i - u| < \epsilon$ for $1 \leq i \leq m$, we have

$$W_1\left(\delta_x Q_{u_1} \cdots Q_{u_m}, \delta_y Q_{u_1} \cdots Q_{u_m}\right) \leq r d(x,y),$$

where m and r are defined in assumption **B2**.

Note. Since the Wasserstein distances satisfy $W_1 \leq W_p$ for $p \geq 1$, if we prove **B4** but using W_p instead of W_1 and if there exists $C_1 \geq 1$ such that for all $(x,y,u) \in E^2 \times [0,1]$, $W_p\left(\delta_x Q_u, \delta_y Q_u\right) \leq C_1 d(x,y)$, then Assumption **B2** will be satisfied. We will mainly proceed like this in our examples.

Proposition 8. *Suppose that Assumptions* **B1** $-$ **B2** *and* **B4** *hold. Then there exist* $C > 0$*, only depending on* m, r, C_1, ϵ *such that*

$$\tau_n(j) \leq C r^{j/m}.$$

Note. Let us remind that Proposition 8 implies a geometric decrease for the covariances. This is a consequence of the following property. If $f : E \to \mathbb{R}$ is measurable and bounded and $g : E \to \mathbb{R}$ is measurable and $\delta(g)-$Lipschitz, we have

$$\operatorname{Cov}\left(f\left(X_{n,i}\right), g\left(X_{n,i+j}\right)\right) \leq \|f\|_\infty \cdot \delta(g) \cdot \tau_n(j).$$

Proof of Proposition 8 We first consider the case $n \geq m/\epsilon$. Now if k is an integer such that $k+m-1 \leq n$, note that assumption **B4** entails that

$$W_1\left(\mu Q_{\frac{k}{n}} \cdots Q_{\frac{k+m-1}{n}}, \nu Q_{\frac{k}{n}} \cdots Q_{\frac{k+m-1}{n}}\right) \leq r W_1(\mu,\nu), \tag{7}$$

where the probability measures μ and ν have both a finite first moment. If $j = mt+s$, we get from (7) and Assumption **B2**,

$$\tau_n(j) \leq C_1^s r^t \sup_{i \in \mathbb{Z}} \mathbb{E}\left[W_1\left(\delta_{X_{n,i}}, \pi_i^{(n)}\right)\right] \leq 2 \sup_{i \in \mathbb{Z}} \mathbb{E} d\left(X_{n,i}, x_0\right) \cdot C_1^s r^t.$$

We have seen in the proof of Theorem 3 that $\sup_{n \in \mathbb{Z}, i \leq n} \mathbb{E} d\left(X_{n,i}, x_0\right) < \infty$.

Now assume that $n < m/\epsilon$. If $j \leq n$, we have

$$\tau_n(j) \leq 2 \sup_{i \in \mathbb{Z}} \mathbb{E} d\left(X_{n,i}, x_0\right) \cdot C_1^{m/\epsilon}.$$

Now if $j > n$, we have since $(X_{n,j})_{j \leq 0}$ is stationary with transition kernel Q_0,

$$\tau_n(j) \leq 2 \sup_{i \in \mathbb{Z}} \mathbb{E} d\left(X_{n,i}, x_0\right) \cdot C_1^{m/\epsilon + m} r^{\left[\frac{j-n}{m}\right]}.$$

This leads to the result for $\rho = r^{1/m}$ and an appropriate choice of $C > 0$.$\square$

8 Poisson GARCH process

Stationary Poisson GARCH processes are widely used for analyzing series of counts. See Fokianos et al. (2009) 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 $Y_k(u)|\sigma\left(Y_{k-j}(u), j\geq 1\right)$ is a Poisson distribution of parameter $\lambda_k(u)$ given recursively by

$$\lambda_k(u)=\gamma(u)+\alpha(u)Y_{k-1}(u)+\beta(u)\lambda_{k-1}(u),$$

where γ,α,β are positive Lipschitz functions such that

$$a=\max_{u\in[0,1]}\left[\alpha(u)+\beta(u)\right]<1.$$

To construct a Markov chain, we consider $X_k(u)=\left(Y_k(u),\lambda_k(u)\right)'$. On $E=\mathbb{R}_+^2$, we consider the distance $d(x,y)=|x_1-y_1|+|x_2-y_2|$ and $x_0=0$, $p=1$. To check **B4**, we consider the following coupling of $\left(\delta_x Q_{u_1}\cdots Q_{u_m},\delta_y Q_{u_1}\cdots Q_{u_m}\right)$ which is also used in Fokianos et al. (2009). Let $N^{(1)},\ldots,N^{(m)}$ be i.i.d Poisson processes of intensity 1. We set $\lambda(0,x)=x_2$, $Y(0,x)=x_1$ and for $1\leq i\leq m$,

$$\lambda(i,x)=\gamma(u_i)+\alpha(u_i)Y(i-1,x)+\beta(u_i)\lambda(i-1,x),$$
$$Y(i,x)=N^{(i)}_{\lambda(i,x)}.$$

The same recursive construction is done with a starting point $y\neq x$. Using independence and stationarity of the increments of a Poisson process, we have

$$\mathbb{E}\left|Y(i,x)-Y(i,y)\right|=\mathbb{E}\left|\lambda(i,x)-\lambda(i,y)\right|.$$

Using this inequality recursively for $1\leq i\leq m$ combined with the triangular inequality, we get

$$\begin{aligned}
& W_1\left(\delta_x Q_{u_1}\cdots Q_{u_m},\delta_y Q_{u_1}\cdots Q_{u_m}\right)\\
\leq\ & \mathbb{E}\left|Y(m,x)-Y(m,y)\right|+\mathbb{E}\left|\lambda(m,x)-\lambda(m,y)\right|\\
\leq\ & 2a^m d(x,y).
\end{aligned}$$

If m is large enough, we have $r=2a^m<1$ which entails **B4**. Assumptions **B2** and **B3** follows in the same way, using our coupling with a Poisson process. Note also that by the Kantorovitch duality, we have for all function $f\in\Delta_1(E)$,

$$\left|\int f d\pi_k^{(n)}-\int f d\pi_u\right|\leq C\left[\left|u-\frac{k}{n}\right|+\frac{1}{n}\right],$$

where $C>0$ is the constant given in Theorem 3. In particular if f only depends on the first coordinate, takes the value 1 at point $j\in\mathbb{N}$ and vanishes outside $[j-0.5,j+0.5]$, we get a bound for the difference $\pi_k^{(n)}(j)-\pi_u(j)$.

9 Extension of the contraction condition in Wasserstein metric for higher order Markov processes

In this section,we give an extension of our result to Markov sequences of order $q\geq 1$ and taking values in the Polish space (E,d). Let $\{S_u : u\in[0,1]\}$ be a family of probability kernels from $(E^q,\mathcal{B}(E^q))$ to $(E,\mathcal{B}(E))$. The two following assumptions will be used.

H1 For all $\mathbf{x} \in E^q$, $S_u(\mathbf{x}, \cdot) \in \mathcal{P}_p(E)$.

H2 There exist non-negative real numbers $a_1, a_2, \ldots, a_q$ satisfying $\sum_{j=1}^q a_j < 1$ and such that for all $(u, \mathbf{x}, \mathbf{y}) \in [0,1] \times E^q \times E^q$,

$$W_p\left(S_u(\mathbf{x}, \cdot), S_u(\mathbf{y}, \cdot)\right) \leq \sum_{j=1}^q a_j d(x_j, y_j).$$

H2 There exists a positive real number C such that for all $(u, , v, \mathbf{x}) \in [0,1] \times [0,1] \times E^q$,

$$W_p\left(S_u(\mathbf{x}, \cdot), S_v(\mathbf{x}, \cdot)\right) \leq C \left(1 + \sum_{j=1}^q d(x_j, x_0)\right) |u - v|.$$

To define Markov chains, we consider the family of Markov kernels $\{Q_u : u \in [0,1]\}$ on the measurable space $(E^q, \mathcal{B}(E)^q)$ and defined by

$$Q_u\left(\mathbf{x}, d\mathbf{y}\right) = S_u\left(\mathbf{x}, dy_q\right) \otimes \delta_{x_2}(y_1) \otimes \cdots \otimes \delta_{x_q}(dy_{q-1}).$$

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

Corollary 3. *If the assumptions* **H1-H3** *hold true then Theorem* 3 *and Proposition* 8 *apply.*

Proof of Corollary 3 Assumption **H1** entails **B1**. Then we check assumption **B3**. If $(u, v, \mathbf{x}) \in [0,1] \times [0,1] \times E^q$, let $\alpha_{\mathbf{x},u,v}$ be a coupling of the two probability distributions $S_u(\mathrm{x}, \cdot)$ and $S_v(\mathrm{x}, \cdot)$. Then

$$\gamma_{\mathbf{x},u,v}(d\mathbf{y}, d\mathbf{y}') = \alpha_{\mathbf{x},u,v}(dy_q, dy_q') \otimes_{j=1}^{q-1} \delta_{x_{j+1}}(dy_j) \otimes \delta_{x_{j+1}}(dy_j')$$

defines a coupling of the two measures $\delta_{\mathbf{x}} Q_u$ and $\delta_{\mathbf{x}} Q_v$. We have

$$W_p\left(\delta_{\mathbf{x}} Q_u, \delta_{\mathbf{x}} Q_v\right) \leq \left[\int d(y_q, y_q')^p \alpha_{\mathbf{x},u,v}(dy_q, dy_q')\right]^{1/p}.$$

By minimizing the last bound over the set of all possible couplings, we get

$$W_p\left(\delta_{\mathbf{x}} Q_u, \delta_{\mathbf{x}} Q_v\right) \leq W_p\left(S_u\left(\mathbf{x}, \cdot\right), S_v\left(\mathbf{x}, \cdot\right)\right),$$

which shows **B3**, using assumption **H3**.
Finally, we check assumptions **B2** and **B4**. For an integer $m \geq 1$, $(u_1, \ldots, u_m) \in [0,1]^m$ and $(\mathbf{x}, \mathbf{y}) \in E^q \times E^q$, we denote by $\alpha_{\mathbf{x},\mathbf{y},u}$ an optimal coupling of $(S_u(\mathbf{x}, \cdot), S_u(\mathbf{y}, \cdot))$. From Villani (2009), Corollary 5.22, there exists a measurable choice of $(\mathbf{x}, \mathbf{y}) \mapsto \alpha_{\mathbf{xy},u}$. We define

$$\gamma_{m,u_1,\ldots,u_m}^{(\mathbf{x_{q+1}} \mathbf{y_{q+1}})}\left(dx_{q+1}, \ldots, dx_{q+m}, dy_{q+1}, \ldots, dy_{q+m}\right) = \prod_{i=q+1}^{q+m} \alpha_{\mathbf{x_i},\mathbf{y_i},u_i}(dx_i, dy_i),$$

where $\mathbf{x_i} = (x_{i-1}, \ldots, x_{i-q})$. Let $\Omega = E^m \times E^m$ endowed with its Borel sigma field and the probability measure $\mathbb{P} = \gamma_{m,u_1,\ldots,u_m}^{(\mathbf{x_{q+1}} \mathbf{y_{q+1}})}$. Then we define the random variables $Z_j^{\mathbf{x_{q+1}}} = x_j$, $Z_j^{\mathbf{y_{q+1}}} = y_j$

for $1 \leq j \leq q$ and for $1 \leq j \leq m$, $Z_{q+j}^{\mathbf{x_{q+1}}}(\omega_1, \omega_2) = \omega_{1,j}$, $Z_{q+j}^{\mathbf{y_{q+1}}}(\omega_1, \omega_2) = \omega_{2,j}$ for $j = 1, \ldots, m$. By definition of our couplings, we have

$$\mathbb{E}^{1/p}\left[d\left(Z_k^{\mathbf{x_{q+1}}}, Z_k^{\mathbf{y_{q+1}}}\right)^p\right] \leq \sum_{j=1}^q a_j \mathbb{E}^{1/p}\left[d\left(Z_{k-j}^{\mathbf{x_{q+1}}}, Z_{k-j}^{\mathbf{y_{q+1}}}\right)^p\right].$$

Using a finite induction, we obtain

$$\mathbb{E}^{1/p}\left[d\left(Z_k^{\mathbf{x_{q+1}}}, Z_k^{\mathbf{y_{q+1}}}\right)^p\right] \leq \alpha^{\frac{k}{q}} \max_{1\leq j\leq q} d(x_j, y_j),$$

where $\alpha = \sum_{j=1}^q a_j$. Setting $X_m^{\mathbf{x}} = \left(Z_{m-q+1}^{\mathbf{x}}, \ldots, Z_m^{\mathbf{x}}\right)$, this entails

$$\begin{aligned}
W_p\left(\delta_{\mathbf{x}} Q_{u_1}\cdots Q_{u_m}, \delta_{\mathbf{y}} Q_{u_1}\cdots Q_{u_m}\right) &\leq \mathbb{E}^{1/p}\left[d_q\left(X_m^{\mathbf{x}}, X_m^{\mathbf{y}}\right)^p\right] \\
&\leq \sum_{j=1}^q \alpha^{\frac{m-j+1}{q}} \max_{1\leq j\leq q} d(x_j, y_j) \\
&\leq \sum_{j=1}^q \alpha^{\frac{m-j+1}{q}} \cdot d_q(\mathbf{x}, \mathbf{y}).
\end{aligned}$$

Then **B2-B4** are satisfied if m is large enough by noticing that $W_1 \leq W_p$. □

Example. Natural examples of $q-$order Markov chains satisfying our assumptions are given by time-varying autoregressive process. More precisely, if E and G are measurable spaces and $F : [0,1] \times E^q \times G \to E$, the triangular array $\{X_{n,i} : 1 \leq i \leq n, n \in \mathbb{Z}^+\}$ is defined recursively by the equations

$$X_{n,i} = F\left(\frac{i}{n}, X_{n,i-1}, \ldots, X_{n,i-q}, \varepsilon_i\right), \quad q+1 \leq i \leq n, \tag{8}$$

where the usual convention is to assume that

$$X_{n,i} = F\left(0, X_{n,i-1}, \ldots, X_{n,i-q}, \varepsilon_i\right), \quad i \leq 0.$$

Then, if $S_u(\mathbf{x}, \cdot)$ denotes the distribution of $F\left(u, x_q, \ldots, x_1, \varepsilon_1\right)$, we have

$$W_p\left(S_u(\mathbf{x}, \cdot), S_u(\mathbf{y}, \cdot)\right) \leq \mathbb{E}^{1/p}\left[d\left(F\left(u, x_q, \ldots, x_1, \varepsilon_1\right), F\left(u, y_q, \ldots, y_1, \varepsilon_1\right)\right)^p\right],$$

$$W_p\left(S_u(\mathbf{x}, \cdot), S_v(\mathbf{x}, \cdot)\right) \leq \mathbb{E}^{1/p}\left[d\left(F\left(u, x_q, \ldots, x_1, \varepsilon_1\right), F\left(v, x_q, \ldots, x_1, \varepsilon_1\right)\right)^p\right].$$

Then the assumptions $\mathbf{H1} - \mathbf{H3}$ are satisfied if for all $(u, v, \mathbf{x}, \mathbf{y}) \in [0,1] \times [0,1] \times E^q \times E^q$,

$$\mathbb{E}^{1/p}\left[d\left(F(u, x_q, \ldots, x_1), x_0\right)^p\right] < \infty,$$

$$\mathbb{E}^{1/p}\left[d\left(F(u, x_q, \ldots, x_1), F(u, y_q, \ldots, y_1)\right)^p\right] \leq \sum_{j=1}^q a_j d\left(x_{q-j+1}, y_{q-j+1}\right)$$

and

$$\mathbb{E}^{1/p}\left[d\left(F(u, x_q, \ldots, x_1), F(v, x_q, \ldots, x_1)\right)^p\right] \leq C\left(1 + \sum_{j=1}^q d(x_j, x_0)\right) \cdot |u - v|.$$

A typical example of such time-varying autoregressive process is the univariate tv-ARCH process for which

$$X_{n,i} = \xi_i \sqrt{a_0(i/n) + \sum_{j=1}^{q} a_j(i/n) X_{n,i-j}^2},$$

with $\mathbb{E}\xi_t = 0$, $\operatorname{Var}\xi_t = 1$. The previous assumptions are satisfied for the square of this process if the a_j's are Lipschitz continuous and if

$$\|\xi_t^2\|_p \cdot \sup_{u\in[0,1]} \sum_{j=1}^{q} a_j(u) < 1, \text{ for some } p \geq 1.$$

See Fryzlewicz et al. (2008) and Truquet (2017a) for the use of those processes for modeling financial data.

Note that one can also consider some autoregressive processes for instance on $\mathbb{R}$ with $d(x,y) = |x-y|^\alpha$, $\alpha \in (0,1)$ and $p = 1$. This is useful to define some models for which the noise ε is not integrable. However, note that in this case, $(\mathbb{R}, d)$ is not a Banach space because d is not associated to a norm.

10 Auxilliary Lemma for the proofs of Section 3

Lemma 2. *If $f : E \to \mathbb{R}$ is a Lipschitz function, then for all measures $\mu, \nu \in \mathcal{P}_p(E)$, we have*

$$\left| \left(\int f^p d\mu \right)^{1/p} - \left(\int f^p d\nu \right)^{1/p} \right| \leq \delta(f) W_p(\mu,\nu),$$

where $\delta(f)$ denotes the Lipschitz constant of f:

$$\delta(f) = \sup_{x\neq y} \frac{|f(x)-f(y)|}{d(x,y)}.$$

Proof of Lemma 2 If γ denotes an optimal coupling for (μ,ν), we get from the triangular inequality,

$$\begin{aligned}
& \left| \left(\int f^p d\mu \right)^{1/p} - \left(\int f^p d\nu \right)^{1/p} \right| \\
= & \left| \left(\int f^p(x) d\gamma(x,y) \right)^{1/p} - \left(\int f^p(y) d\gamma(x,y) \right)^{1/p} \right| \\
\leq & \left(\int |f(x)-f(y)|^p d\gamma(x,y) \right)^{1/p} \\
\leq & \delta(f) \left(\int d(x,y)^p d\gamma(x,y) \right)^{1/p}.
\end{aligned}$$

which leads to the result of the lemma.□

Lemma 3. *Let X and Y two random variables taking values in (E, d) and such that $\mathbb{P}_X$, $\mathbb{P}_Y \in \mathcal{P}_d(E)$. On $E \times E$, we define the metric*

$$\widetilde{d}\left((x_1, x_2), (y_1, y_2)\right) = \left(d(x_1, y_1)^p + d(x_2, y_2)^p\right)^{1/p}.$$

Then we have

$$W_p\left(\mathbb{P}_{X,Y}, \mathbb{P}_{Y,Y}\right) \geq 2^{-\frac{p-1}{p}} \mathbb{E}^{1/p}\left(d(X, Y)^p\right).$$

Proof of Lemma 3 Consider the Lipschitz function $f : E \times E \to \mathbb{R}$ defined by $f(x_1, x_2) = d(x_1, x_2)$. Using the triangular inequality and convexity, we have $\delta(f) \leq 2^{\frac{p-1}{p}}$. Then the result is a consequence of Lemma 2.$\square$

Lemma 4. *Let $\mu \in \mathcal{P}_p(E)$ and Q, R be two probability kernels from $(E, \mathcal{B}(E))$ to $(E, \mathcal{B}(E))$ such that*

1. *for all $x \in E$, the two probability measures $\delta_x Q$ and $\delta_x R$ are elements of $\mathcal{P}_p(E)$,*
2. *there exists $C > 0$ such that for all $(x, y) \in E^2$,*

$$W_p\left(\delta_x Q, \delta_y Q\right) \leq C d(x, y), \quad W_p\left(\delta_x R, \delta_y R\right) \leq C d(x, y).$$

Then, if $\mu \in \mathcal{P}_p(E)$, the two probability measures μQ, μR are also elements of $\mathcal{P}_p(E)$. Moreover, we have

$$W_p^p\left(\mu Q, \mu R\right) \leq \int W_p^p\left(\delta_x Q, \delta_x R\right) d\mu(x), \tag{9}$$

and if ν is another element of $\mathcal{P}_p(E)$, we have

$$W_p\left(\mu Q, \nu Q\right) \leq C W_p(\mu, \nu). \tag{10}$$

Proof of Lemma 4. Using Lemma 3 with $f(x) = d(x, x_0)$, we have for a given $y \in E$,

$$\begin{aligned}
\int d(x, x_0)^p Q(y, dx) &\leq \left[W_p(\delta_y Q, \delta_{x_0} Q) + \left(\int d(x, x_0)^p Q(x_0, dx)\right)^{1/p}\right]^p \\
&\leq \left[C d(x_0, y) + \left(\int d(x, x_0)^p Q(x_0, dx)\right)^{1/p}\right]^p.
\end{aligned}$$

After integration with respect to μ, it is easily seen that $\mu Q \in \mathcal{P}_p(E)$.
To show (9), one can use Kantorovitch duality (see Villani (2009), Theorem 5.10). Denoting by $\mathcal{C}_b(E)$ the set of bounded continuous functions on E, we have

$$\begin{aligned}
W_p^p\left(\mu Q, \mu R\right) &= \sup_{\phi(x)-\psi(y)\leq d(x,y)^p, (\phi,\psi)\in\mathcal{C}_b(E)} \left\{\int \phi(x)\mu Q(dx) - \int \psi(y)\mu R(dy)\right\} \\
&\leq \int \left[\sup_{\phi(x)-\psi(y)\leq d(x,y)^p, (\phi,\psi)\in\mathcal{C}_b(E)} \left\{\int \phi(x) Q(z, dx) - \int \psi(y) R(z, dy)\right\}\right] \mu(dz) \\
&\leq \int W_p^p\left(\delta_z Q, \delta_z R\right) \mu(dz).
\end{aligned}$$

Finally, we show (10). Let ϕ, ψ be two elements of $\mathcal{C}_b(E)$ such that $\phi(x) - \psi(y) \leq d(x,y)^p$ and γ an optimal coupling for (μ, ν). Then, for $u, v \in E$, we have

$$\int \phi(x) Q(u, dx) - \int \psi(y) Q(v, dy) \leq W_p^p(\delta_u Q, \delta_v Q) \leq C^p d(u,v)^p.$$

Moreover,

$$\int \phi(x) \mu Q(dx) - \int \psi(y) \nu Q(dy) = \int \gamma(du, dv) \left[\int \phi(x) Q(u, dx) - \int \psi(y) Q(v, dy) \right].$$

Then (10) easily follows from Kantorovitch duality.□

Lemma 5. *Let $j \geq 1$ be an integer. Assume that $Q_1, \ldots, Q_j$ and $R_1, \ldots, R_j$ are Markov kernels such that for all $x \in E$ and $1 \leq i \leq j$, $\delta_x Q_i$ and $\delta_x R_i$ are elements of $\mathcal{P}_p(E)$ satisfying*

$$W_p(\delta_x Q_i, \delta_y Q_i) \leq L_i d(x,y), \quad W_p(\delta_x R_i, \delta_y R_i) \leq L_i d(x,y),$$

for all $(x,y) \in E^2$. Then, for all $x \in E$, we have

$$W_p(\delta_x Q_1 \cdots Q_j, \delta_x R_1 \cdots R_j) \leq \sum_{s=0}^{j-1} L_j \cdots L_{j-s+1} D_{j-s},$$

where $D_i^p = \int W_p^p(\delta_y Q_i, \delta_y R_i) \delta_x R_1 \cdots R_{i-1}(dy)$.

Proof of Lemma 5 Using the inequality

$$W_p(\delta_x Q_1 \cdots Q_j, \delta_x R_1 \cdots R_j) \leq \sum_{s=0}^{j-1} W_p(\delta_x R_1 \cdots R_{j-s-1} Q_{j-s} \cdots Q_j, \delta_x R_1 \cdots R_{j-s} Q_{j-s+1} \cdots Q_j),$$

the result follows using Lemma 4.□

11 Additional examples for the Wasserstein metric

11.1 Iterated random affine functions

In this section, we consider some examples of iteration of random affine functions. Here we assume that for each $u \in [0,1]$, there exists a sequence $(A_t(u), B_t(u))_{t \in \mathbb{Z}}$ of i.i.d random variables such that $A_t(u)$ takes its values in the space $\mathcal{M}_d$ of squares matrices of dimension d with real coefficients and $B_t(u)$ takes its values in $E = \mathbb{R}^d$. Let $\|\cdot\|$ a norm on E. We also denote by $\|\cdot\|$ the corresponding operator norm on $\mathcal{M}_d$. We then consider the following recursive equations

$$X_{n,i} = A_i\left(\frac{i}{n}\right) X_{n,i-1} + B_i\left(\frac{i}{n}\right). \tag{11}$$

Local approximation of these autoregressive processes by their stationary versions $X_t(u) = A_t(u) X_{t-1}(u) + B_t(u)$ is studied is studied by Subba Rao (2006). In this subsection, we will derive similar results using our Markov chain approach. For each $u \in [0,1]$, we denote by γ_u the top Lyapunov exponent of the sequence $(A_t(u))_{t \in \mathbb{Z}}$, i.e

$$\gamma_u = \inf_{n \geq 1} \frac{1}{n} \mathbb{E} \log \|A_n(u) A_{n-1}(u) \cdots A_1(u)\|.$$

We assume that there exists $t \in (0,1)$ such that

R1 for all $u \in [0,1]$, $\mathbb{E}\|A_1(u)\|^t < \infty$, $\mathbb{E}\|B_1(u)\|^t < \infty$ and $\gamma_u < 0$.

R2 There exists $C > 0$ such that for all $(u,v) \in [0,1]^2$,

$$\mathbb{E}\|A_1(u) - A_1(v)\|^t + \mathbb{E}\|B_1(u) - B_1(v)\|^t \leq C|u-v|^t.$$

As pointed out in the main document, our results are valid if we assume Hölder continuity instead of Lipschitz continuity, in particular Theorem 3. Then assuming this extension, we get the following result.

Proposition 9. *For $s \in (0,1)$, we set $d(x,y) = \|x-y\|^s$ and $x_0 = 0$. Assume that assumptions* **R1** $-$ **R2** *hold true. Then there exists $s \in (0,t)$ such that for all integer j, there exists a real number $C > 0$ such that for all $u \in [0,1]$ and $1 \leq k \leq n-j+1$,*

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

Notes

1. Using the remark given in the Note of Section 3.2 of the main document, we also have $\mathbb{E}\|X_{n,k} - X_k(u)\|^s \leq C\left(\left|u - \frac{k}{n}\right|^s + \frac{1}{n^s}\right)$, where the process $(X_j(u))_{j\in\mathbb{Z}}$ satisfies the iterations $X_k(u) = A_k(u)X_{k-1}(u) + B_k(u)$. Then the triangular array $\{X_{n,k} : k \leq n, n \in \mathbb{Z}^+\}$ is locally stationary in the sense given in Vogt (2012) (see Definition 2.1 of that paper).

2. One can also give additional results for the Wasserstein metric of order $p \geq 1$ and $d(x,y) = \|x-y\|$ if

   $$\mathbb{E}\|A_1(u)\|^p + \mathbb{E}\|B_1(u)\|^p < \infty, \quad \mathbb{E}^{1/p}\|A_1(u) - A_1(v)\|^p + \mathbb{E}^{1/p}\|B_1(u) - B_1(v)\|^p \leq C|u-v|$$

   and there exists an integer $m \geq 1$ such that $\sup_{u\in[0,1]} \mathbb{E}\|A_m(u)\cdots A_1(u)\|^p < 1$. In particular, one can recover results about the local approximation of tv-AR processes defined by

   $$X_{n,i} = \sum_{j=1}^{q} a_j(i/n)X_{n,i-j} + \sigma(i/n)\varepsilon_i$$

   by vectorizing q successive coordinates and assuming Lipschitz continuity for the a_j's and σ. Details are omitted.

Proof of Proposition 9 For all $(x,u) \in \mathbb{R}^d \times [0,1]$, the measure $\delta_x Q_u$ is the probability distribution of the random variable $A_k(u)x + B_k(u)$. Condition **B1** follows directly from assumption **R1** (whatever the value of $s \in (0,t)$). Moreover, we have for $s \in (0,t)$,

$$\begin{aligned} W_1(\delta_x Q_u, \delta_x Q_v) &\leq \mathbb{E}\|A_k(u) - A_k(v)\|^s \cdot \|x\|^s + \mathbb{E}\|B_k(u) - B_k(v)\|^s \\ &\leq (1 + \|x\|^s) \cdot \left(\mathbb{E}^{\frac{s}{t}}\|A_k(u) - A_k(v)\|^t + \mathbb{E}^{\frac{s}{t}}\|B_k(u) - B_k(v)\|^t\right). \end{aligned}$$

This entails condition **B3**, using assumption **R2**. Next, if $u \in [0,1]$, the conditions $\gamma_u < 0$ and $\mathbb{E}\|A_t(u)\|^t < \infty$ entail the existence of an integer k_u and $s_u \in (0,t)$ such that $\mathbb{E}\|A_{k_u}(u)A_{k_u-1}(u)\cdots A_1(u)\|^{s_u} <$

1 (see for instance Francq and Zakoïan (2010), Lemma 2.3). Using the axiom of choice, let us select for each u, a couple (k_u, s_u) satisfying the previous property. From assumption **R2**, the set

$$\mathcal{O}_u = \{v \in [0,1] : \mathbb{E}\|A_{k_u}(v)A_{k_u-1}(v)\cdots A_1(v)\|^{s_u} < 1\}$$

is an open set of $[0,1]$. By a compactness argument, there exist $u_1,\ldots,u_d \in [0,1]$ such that $[0,1] = \cup_{i=1}^d \mathcal{O}_{u_i}$. Then setting $s = \min_{1\leq i\leq d} s_{u_i}$ and denoting by m the lowest common multiple of the integers $k_{u_1},\ldots,k_{u_d}$, we have from assumption **R2**,

$$r = \sup_{u\in[0,1]} \mathbb{E}\|A_m(u)\cdots A_1(u)\|^s < 1.$$

This entails condition **B2** for this choice of s, m and r. Indeed, we have

$$W_1\left(\delta_x Q_u^m, \delta_y Q_u^m\right) \leq \mathbb{E}\|A_m(u)\cdots A_1(u)(x-y)\|^s \leq r d(x,y).$$

Note also that condition **B4** easily follows from the uniform continuity of the application $(u_1,\ldots,u_m) \mapsto \mathbb{E}\|A_m(u_1)\cdots A_1(u_m)\|^s$.$\square$.

11.2 Additional discussion

The approximation of time-varying autoregressive processes by stationnary processes is discussed in several papers. See for instance Subba Rao (2006) for linear autoregressions with time varying random coefficients, Vogt (2012) for nonlinear time-varying autoregressions or Zhang and Wu (2015) for additional results in the same setting. In what follows, we assume $p=1$ for simplicity. The approximating stationary process of (8) is given by

$$X_i(u) = F\left(u, X_{i-1}(u), \varepsilon_i\right).$$

Note that $W_p\left(\pi_k^{(n)}, \pi_u\right) \leq \mathbb{E}^{1/p}\left[d\left(X_k^{(n)}, X_k(u)\right)^p\right]$ and the aforementioned references usually study a control of this upper bound by $\left|u - \frac{k}{n}\right| + \frac{1}{n}$. Note that in the case of autoregressive processes, a coupling of the time-varying processes and its stationary approximation is already defined because the same noise process is used in both cases. However it is possible to construct some examples for which $\pi_k^{(n)} = \pi_u$ and $\mathbb{E}^{1/p}\left[d\left(X_{n,k}, X_k(u)\right)^p\right] \neq 0$, i.e the coupling used is not optimal. Nevertheless, it is still possible to obtain an upper bound of $\mathbb{E}^{1/p}\left[d\left(X_k^{(n)}, X_k(u)\right)^p\right]$ using our results. To this end, let us consider the Markov kernel form $\left(E^2, \mathcal{B}(E^2)\right)$ to itself and given by

$$Q_v^{(u)}\left(x_1, x_2, A\right) = \mathbb{P}\left(\left(F(v, x_1, \varepsilon_1), F(u, x_2, \varepsilon_1)\right) \in A\right), \quad A \in \mathcal{B}(E^2), \quad v \in [0,1].$$

One can show that the family $\left\{Q_v^{(u)} : v \in [0,1]\right\}$ satisfies the assumptions **B1** $-$ **B3** for the metric

$$d_2\left[(x_1,x_2),(y_1,y_2)\right] = \left(d(x_1,y_1)^p + d(x_2,y_2)^p\right)^{1/p}.$$

Moreover, the proof of Theorem 3 shows that the constant $C>0$ does not depend on $u \in [0,1]$. Then Lemma 3 guarantees that there exists a positive constant C not depending on k,n,u such that

$$\mathbb{E}^{1/p}\left(d\left(X_{k,n}, X_k(u)\right)^p\right) \leq C\left[\left|u - \frac{k}{n}\right| + \frac{1}{n}\right].$$

12 Proof of Proposition 2

1. According to Lemma 1 given in the paper, there exists $(\gamma, \delta) \in (0,1)^2$ only depending λ, b, η, such that

$$\Delta_{V_\delta}(Q_u^m) = \sup\left\{\frac{\|\mu Q_u^m - \nu Q_u^m\|_{V_\delta}}{\|\mu - \nu\|_{V_\delta}} : \quad \mu, \nu \in \mathcal{P}(E), \quad \mu V_\delta < \infty, \nu V_\delta < \infty\right\} \leq \gamma,$$

with $V_\delta = 1 - \delta + \delta V$. From Theorem 6.19 in Douc et al. (2014) and Assumptions $\mathbf{F1} - \mathbf{F2}$, we have a unique invariant probability π_u for Q_u, satisfying $\pi_u V < \infty$ and for $\mu \in \mathcal{P}(E)$ such that $\mu V < \infty$, we have

$$\|\mu Q_u^j - \pi_u\|_{V_\delta} \leq \max_{0 \leq s \leq m-1} \Delta_{V_\delta}(Q_u^s)\gamma^{[j/m]}\|\mu - \pi_u\|_{V_\delta}.$$

Note that $\|\cdot\|_{V_\delta} \leq \|\cdot\|_V \leq \frac{1}{\delta}\|\cdot\|_{V_\delta}$ and the two norms are equivalent. Using Lemma 6.18 in Douc et al. (2014), we have

$$\Delta_{V_\delta}(Q_u^s) = \sup_{x \neq y} \frac{\|\delta_x Q_u^s - \delta_y Q_u^s\|_{V_\delta}}{V_\delta(x) + V_\delta(y)} \leq \frac{K^s}{\delta}.$$

Then it remains to show that $\sup_{u \in [0,1]} \pi_u V < \infty$ or equivalently $\sup_{u \in [0,1]} \pi_u V_\delta < \infty$. But this a consequence of the contraction property of the application $\mu \mapsto \mu Q_u^m$ on the space

$$\mathcal{M}_\delta = \{\mu \in \mathcal{P}(E) : \mu V_\delta < \infty\}$$

endowed with the distance $d_\delta(\mu, \nu) = \|\mu - \nu\|_{V_\delta}$, which is a complete metric space (see Proposition 6.16 in Douc et al. (2014)). Hence we have

$$\mu - \pi_u = \sum_{j=0}^{\infty}\left[\mu Q_u^{mj} - \mu Q_u^{m(j+1)}\right]$$

which defines a normally convergent series in $\mathcal{M}_\delta$ and

$$\|\mu - \pi_u\|_{V_\delta} \leq \sum_{j=0}^{\infty}\gamma^j\|\mu - \mu Q_u^m\|_{V_\delta} \leq \frac{\mu V + K^m \mu V}{1 - \gamma}.$$

This shows that $\sup_{u \in [0,1]} \pi_u V < \infty$ and the proof of the first point is now complete.

2. To prove the second point, we use the decomposition $\pi_u - \pi_v = \pi_u Q_u^m - \pi_v Q_u^m + \pi_v Q_u^m - \pi_v Q_v^m$. This leads to the inequality

$$\|\pi_u - \pi_v\|_{V_\delta} \leq \frac{\|\pi_v Q_u^m - \pi_v Q_v^m\|_{V_\delta}}{1 - \gamma}.$$

Moreover, we have

$$\begin{aligned}\|\pi_v Q_u^m - \pi_v Q_v^m\|_{V_\delta} &\leq \|\pi_v Q_u^m - \pi_v Q_v^m\|_V \\ &\leq \sum_{j=0}^{m-1} \|\pi_v Q_v^{m-j-1}(Q_v - Q_u)Q_u^j\|_V \\ &\leq \sum_{j=0}^{m-1} K^j \|\pi_v Q_v^{m-j-1}(Q_v - Q_u)\|_V \\ &\leq \sum_{j=0}^{m-1} K^j \cdot \pi_v \widetilde{V} \cdot |u - v|.\end{aligned}$$

Hence the result follows with $C = \frac{\sup_{u\in[0,1]} \pi_u \widetilde{V}}{\delta(1-\gamma)} \sum_{j=0}^{m-1} K^j$.□

13 Proof Theorem 3

1. We start with the case $j = 1$. We assume first that $n \geq \frac{m}{\epsilon}$. Under the assumptions of the theorem, Lemma 6 given below guarantees the existence of $(\gamma, \delta) \in (0,1)^2$ such that for all $k \leq n$, $\Delta_{V_\delta}\left(Q_{\frac{k-m+1}{n}} \cdots Q_{\frac{k}{n}}\right) \leq \gamma$ with $V_\delta = 1 - \delta + \delta V$. Set $R_{k,m} = Q_{\frac{k-m+1}{n}} \cdots Q_{\frac{k}{n}}$, we get
$$\|\pi_k^{(n)} - \pi_u\|_{V_\delta} \leq \gamma \|\pi_{k-m}^{(n)} - \pi_u\|_{V_\delta} + \|\pi_u R_{k,m} - \pi_u P_u^m\|_{V_\delta}.$$
From our assumptions, we have
$$\begin{aligned}\|\pi_u R_{k,m} - \pi_u Q_u^m\|_V &\leq \sum_{j=0}^{m-1} \|\pi_u Q_u^{m-j-1}\left[Q_{\frac{k-j}{n}} - Q_u\right] Q_{\frac{k-j+1}{n}} \cdots Q_{\frac{k}{n}}\|_V \\ &\leq \sum_{j=0}^{m-1} K^j \cdot \pi_u Q_u^{m-j+1}\widetilde{V} \cdot \left|u - \frac{k-j}{n}\right| \\ &\leq K^m \sup_{u\in[0,1]} \pi_u \widetilde{V} \sum_{s=k-m+1}^{k} \left|u - \frac{s}{n}\right| \\ &\leq D_1 \left[\left|u - \frac{k}{n}\right| + \frac{1}{n}\right],\end{aligned}$$
for a suitable constant $D_1 > 0$. Noticing that $\|\cdot\|_{V_\delta} \leq \|\cdot\|_V \leq \delta^{-1}\|\cdot\|_{V_\delta}$, we get for a suitable constant $D_2 > 0$,
$$\|\pi_k^{(n)} - \pi_u\|_V \leq \frac{D_1}{\delta} \sum_{j\geq 0} \gamma^j \left[\left|u - \frac{k-jm}{n}\right| + \frac{1}{n}\right] \leq D_2 \left[\left|u - \frac{k}{n}\right| + \frac{1}{n}\right].$$
For $n < m/\epsilon$, $\|\pi_k^n - \pi_u\|_V \leq \pi_k^{(n)} V + \sup_{u\in[0,1]} \pi_u V \leq 2K^{\frac{m}{\epsilon}} \sup_{u\in[0,1]} \pi_u V \cdot \frac{m}{\epsilon} n^{-1}$. Setting $C = \max\left\{D_2, 2K^{\frac{m}{\epsilon}} \frac{m}{\epsilon} \sup_{u\in[0,1]} \pi_u V\right\}$, we get the result.

2. Assume that the result is true for an integer $j \geq 1$. Let $f : E^{j+1} \to \mathbb{R}_+$ be such that $f(x_1, \ldots, x_{j+1}) \leq V(x_1) + \cdots + V(x_{j+1})$. Setting $s_j = \sum_{i=1}^j V(x_i)$ and $g_j(x_{j+1}) = f(x_1, \ldots, x_{j+1})$, we use the decomposition $g_j = g_j \mathbb{1}_{g_j \leq V} + (g_j - V)\, \mathbb{1}_{V < g_j \leq V + s_j} + V \mathbb{1}_{V < g_j \leq V + s_j}$, and we get

$$
\begin{aligned}
& \left| \delta_{x_j} Q_{\frac{k+j+1}{n}} g_j - \delta_{x_{j+1}} Q_u g_j \right| \\
\leq\ & 2 \left\| \delta_{x_j} Q_{\frac{k+j+1}{n}} - \delta_{x_j} Q_u \right\|_V + s_j \| \delta_{x_j} Q_{\frac{k+j+1}{n}} - \delta_{x_j} Q_u \|_{TV} \\
\leq\ & 2 \left[\widetilde{V}(x_j) + (V(x_1) + \cdots + V(x_j))\, L(x_j) \right] \left| u - \frac{k+j+1}{n} \right|,
\end{aligned}
$$

and $\|\pi_{u,j} \otimes Q_{\frac{k+j+1}{n}} - \pi_{u,j+1}\|_V \leq 2 \left(\sup_{u \in [0,1]} \pi_u \widetilde{V} + jG \right) \left| u - \frac{k+j+1}{n} \right|$. Moreover, $\|\pi_{u,j} \otimes Q_{\frac{k+j+1}{n}} - \pi^{(n)}_{k,j+1}\|_V \leq (1 + K) \|\pi^{(n)}_{k,j} - \pi_{u,j}\|_V$. The last two bounds lead to the result by induction. Moreover, using the same type of arguments, one can also check the continuity condition 1 of Definition 1.$\square$

The following result is proved in Hairer and Mattingly (2011), Theorem 1.3. Note that our function V corresponds to their function $V + 1$. For a Markov kernel P on $(E, \mathcal{B}(E))$, we set

$$
\Delta_V(P) = \sup \left\{ \frac{\|\mu P - \nu P\|_V}{\|\mu - \nu\|_V} : \quad \mu, \nu \in \mathcal{P}(E), \quad \mu V < \infty, \nu V < \infty \right\}.
$$

Lemma 6. *Under the assumptions* $\mathbf{F1} - \mathbf{F2}$*, there exists* $(\gamma, \delta) \in (0,1)^2$*, only depending on* λ, η, b*, such that for all* $(u_1, \ldots, u_m) \in [0,1]^m$ *such that* $|u_i - u_j| \leq \epsilon$*,* $1 \leq i, j \leq m$*, we have* $\Delta_{V_\delta}(Q_{u_1} \cdots Q_{u_m}) \leq \gamma$*, with* $V_\delta = 1 - \delta + \delta V$.

14 Mixing properties of ARCH processes

We assume that

$$
X_{n,k} = \varepsilon_k \sqrt{a_0(k/n) + \sum_{j=1}^p a_j(k/n) X_{n,k-j}^2} = \varepsilon_k \sigma_{k/n}(X_{n,k-1}, \ldots, X_{n,k-p}).
$$

with continuous coefficients $a_0, \ldots, a_p$, a positive function a_0, $\alpha = \sup_{u \in [0,1]} \sum_{j=1}^p a_j(u) < 1$ and the noise has a density f_ϵ which has a positive lower-bound on each compact set. Set $c = \max_{u \in [0,1]} a_0(u) + 1 - \alpha$. Here we have

$$
Q_u(\mathbf{x}, d\mathbf{y}) = \frac{1}{\sigma_u(y_{p-1}, x_{p-1}, \ldots, x_2)} f_\varepsilon \left(\frac{y_p}{\sigma_u(y_{p-1}, x_{p-1}, \ldots, x_2)} \right) \prod_{i=1}^{p-1} \delta_{x_{i+1}} dy_i.
$$

Now set $V_1(x) = 1 + x^2$. For $m \geq 1$ and $u_1, \ldots, u_m \in [0,1]$, we define the recursion

$$
Y_i = \varepsilon_i \sigma_{u_i}(Y_{i-1}, \ldots, Y_{i-p}), \quad Y_1 = x_1, \ldots, Y_p = x_p, \quad i = p+1, \ldots, p+m.
$$

For $p + 1 \leq i \leq p + m$, we have $d_i = \mathbb{E} V_1(Y_i) \leq c + \alpha \max\{d_{i-1}, \ldots, d_{i-p}\}$. Setting $V(\mathbf{x}) = \sum_{i=1}^p V_1(x_i)$, we have $\delta_{\mathbf{x}} Q_{u_m} \cdots Q_{u_1} V = \sum_{i=0}^{p-1} \mathbb{E} V_1(Y_{m+p-i})$. By a finite induction we obtain

$$
d_i \leq \frac{c}{1 - \alpha} + \alpha^{\frac{i-p-1}{p}} V_p(\mathbf{x}), \quad p + 1 \leq i \leq p + m.
$$

Then **F1** is satisfied if m is large enough. In the sequel, we assume that $m \geq p$. Next we set

$$g_u(x_1, \ldots, x_{p+1}) = \frac{1}{\sigma_u(x_p, \ldots, x_1)} f_\varepsilon \left(\frac{x_{p+1}}{\sigma_u(x_p, \ldots, x_1)} \right)$$

and for $a \leq b$,

$$s = \inf \left\{ g_u(x_1, \ldots, x_{p+1}) : u \in [0,1], (x_1, \ldots, x_{p+1}) \in [a,b]^{p+1} \right\}.$$

If $x_1, \ldots, x_p \in [a,b]$, we have

$$\delta_x Q_{u_m} \cdots Q_{u_1}(A) \geq s^m (b-a)^{m-p} \lambda(A \cap [a,b]^p),$$

where λ denotes the Lebesgue measure on $\mathbb{R}^p$. We first choose $R > 2b/(1-\lambda)$ and then $a \leq b$ such that $\{V \leq R\} \subset [a,b]^p$, assumption **F2** follows with ν being the uniform measure on $[a,b]^p$.

Note. Our assumption for the density f_ε is different from that used in Fryzlewicz and Subba Rao (2011). Indeed Fryzlewicz and Subba Rao (2011) use (see Assumption 3.1) a continuity type condition while we impose a positivity condition.

15 Justifications for the example of Section 4: Markov switching autoregressive processes

We will use the following additional assumptions.

1. There exists a positive integer p such that

$$\limsup_{y \to \infty} \max_{z \in E_2} \sup_{u \in [0,1]} \sum_{z' \in E_2} \overline{Q}_u(z,z') \frac{\mathbb{E}\left| m(u,z',y) + \sigma(u,z',y)\varepsilon_1 \right|^{p+1}}{|y|^{p+1}} < 1.$$

2. We assume that $\inf_{(u,y,z) \in [0,1] \times E} \sigma(u,z,y) > 0$. Moreover, there exists a positive constant C such that for all $(u,v,y,z) \in [0,1]^2 \times E$,

$$|m(u,z,y) - m(v,z,y)| + |\sigma(u,z,y) - \sigma(v,z,y)| \leq C(1+|y|)|u-v|, \quad |m(0,z,y)| + \sigma(0,z,y) \leq C(1+|y|).$$

3. The noise density f_ε is positive everywhere, continuously differentiable and satisfies $\int |z|^{p+1} f_\varepsilon(z) dz < \infty$ and $\int |z|^{p+1} |f_\varepsilon'(s)| dz < \infty$.

The transition kernel Q_u for the bivariate Markov chain $X_k(u) = (Y_k(u), Z_k(u))'$ is defined by

$$\delta_{y_1,z_1} Q_u(A \times \{z_2\}) = \overline{Q}_u(z_1,z_2) \int_A \frac{1}{\sigma(u,z_2,y_1)} f_\varepsilon \left(\frac{y_2 - m(u,z_2,y_1)}{\sigma(u,z_2,y_1)} \right) dy_2.$$

Let us show that Assumptions **F1-F3** are satisfied.

- To check the small set condition **F2**, we first choose an integer $m \geq 1$ such that $\inf_{u\in[0,1]} Q_u^m(z,z') > 0$ for all $(z,z') \in E_2^2$. Using uniform continuity, we have $\delta_{z,z'} = \inf_{(u_1,\ldots,u_m)\in I_m(\epsilon)} Q_{u_1}\cdots Q_{u_m}(z,z') > 0$ for all $(z,z') \in E_2^2$, provided $\epsilon > 0$ is large enough. We will show that the small set condition is satisfied for all compact sets. Let K be a compact set. We set $g(u,z,y,y') = \frac{1}{\sigma(u,z,y)} f_\varepsilon\left(\frac{y'-m(u,z,y)}{\sigma(u,z,y)}\right)$. If $y_0 \in K$ and $z_0 \in E_2$, we have for a Borel set A,

$$\begin{aligned} \delta_{y_0,z_0} Q_{u_1}\cdots Q_{u_m}(A\times\{z\}) &\geq \min_{z,z'\in E_2} \delta_{z,z'} \cdot \left(\inf_{(u,z,y,y')\in[0,1]\times E_2\times[0,1]^2} g(u,z,y,y')\right)^{m-1} \\ &\times \inf_{(u,z,y,y')\in[0,1]\times E_2\times K\times[0,1]} g(u,z,y,y')\cdot\nu(A) \end{aligned}$$

  where ν denotes the uniform measure over $[0,1]$. From our assumptions, the noise density has a positive lower bound on each compact subset of $\mathbb{R}$ and the functions m and σ are locally bounded. This entails

$$\left(\inf_{(u,z,y,y')\in[0,1]\times E_2\times[0,1]^2} g(u,z,y,y')\right)^{m-1} \cdot \inf_{(u,z,y,y')\in[0,1]\times E_2\times K\times[0,1]} g(u,z,y,y') > 0.$$

  The small set condition is then satisfied for all compact subsets of E.

- Now, we check **F1**. We set $\widetilde{V}(y,z) = 1 + |y|^{p+1}$ and $V(y,z) = 1+|y|^p$. From our first assumption, it is clear that there exist $\lambda \in (0,1)$ and $b > 0$ such that for all $(u,z,y) \in [0,1]\times E_2\times\mathbb{R}$,

$$\delta_{y,z} Q_u \widetilde{V} \leq \lambda V(y,z) + b.$$

  The result also holds for V instead of $\widetilde{V}$. Iterating this inequality, we see that **F1** is also satisfied by iterating m Markov kernels where m is defined in the previous point. Note also that the set $\{V \leq R\}$ is compact for all values of R and the **F2** is also satisfied whatever the value of $R > 0$. Moreover, from the drift condition satisfied by $\widetilde{V}$, we have $\sup_{u\in[0,1]} \pi_u \widetilde{V} < \infty$, a property necessary to check Assumption **F3**.

- Using our second assumption, the moment condition for the noise and the Lipschitz properties of $u \mapsto \overline{Q}_u$, there exists a positive constant $\overline{C}$ such that

$$\begin{aligned} \|\delta_{y,z}Q_u - \delta_{y,z}Q_v\|_V &\leq \sum_{z'\in E_2} \overline{Q}_u(z,z') \int \left[1+|y'|^p\right]\cdot\left|g(u,z',y,y') - g(v,z',y,y')\right| dy' \\ &+ \overline{C}|u-v|\cdot(1+|y|^p) \\ &= E_1 + E_2. \end{aligned}$$

  For simplicity of notations, the quantities $m(u,z,y)$ and $\sigma(u,z,y)$ will be simply denoted by m and σ respectively and $m(v,z,y), \sigma(v,z,y)$ by $m'$and σ'. We have the bound $E_1 \leq A + B$ with

$$A = \int \left[1+|y'|^p\right] \frac{|\sigma-\sigma'|}{\sigma\sigma'} f_\varepsilon\left(\frac{y'-m'}{\sigma'}\right) dy'$$

  and

$$B = \int \left[1+|y'|^p\right] \frac{1}{\sigma} \left| f_\varepsilon\left(\frac{y'-m}{\sigma}\right) - f_\varepsilon\left(\frac{y'-m'}{\sigma'}\right)\right| dy'.$$

Using our assumptions and the mean value theorem for bounding B, we have

$$A \leq L\left(1+|y|^{p+1}\right)\cdot\left(1+\int |w|^p f_\varepsilon(w)dw\right)$$

and

$$B \leq L\left(1+|y|^{p+1}\right)\cdot\left(1+\int |w|^{p+1}\left|f'_\varepsilon(w)\right|dw\right)$$

for a suitably chosen constant $L>0$. This shows **F3**. Finally the function L in the second point of Theorem 3 is of the form (using the previous bounds for $p=0$) $L(y)=K(1+|y|)$ and the integrability condition follows form $\sup_{u\in[0,1]}\pi_u\widetilde{V}<\infty$.

Note If the functions m and σ do not depend on their first argument, the assumptions given above can be weakened. In Assumption 2, it is only necessary to assume a positive lower bound for σ and that for all compact subset K of $\mathbb{R}$,

$$\sup_{y\in K}\max_{z\in E_2}\left[|m(z,y)|+\sigma(z,y)\right]<\infty.$$

Assumption 3 can be replaced by: the noise density f_ε has a positive lower bound on each compact subset of $\mathbb{R}$. Finally one can take $V(y,z)=\widetilde{V}(y,z)=1+|y|^p$, p being a positive integer, for checking Assumption **F3**.

16 Justifications for the local stationarity of integer-valued autoregressive processes

We consider a sequence $(Y_i(u))_{i\geq 0}$ of i.i.d random variables following the Bernoulli (resp. Poisson for the Poisson ARCH process) distribution of parameter α_u and a random variable $\xi(u)$ following the Poisson distribution of parameter λ_u. We assume that $\xi(u)$ and the sequence $(Y_i(u))_{i\geq 0}$ are independent. For $u\in[0,1]$, we have

$$\begin{aligned}
&\delta_x Q_u V_p\\
=\;& 1+\mathbb{E}\left(\alpha_u x+\sum_{i=1}^{x}\left(Y_i(u)-\alpha_u\right)+\xi(u)\right)^p\\
=\;& 1+\alpha_u^p x^p+\sum_{j=1}^{p}\binom{p}{j}\alpha_u^{p-j}x^{p-j}\mathbb{E}\left(\sum_{i=1}^{x}\left(Y_i(u)-\alpha_u\right)+\xi(u)\right)^j
\end{aligned}$$

Using the Burkhölder inequality for martingales, we have for an integer $\ell\geq 2$,

$$\mathbb{E}\left(\sum_{i=1}^{x}\left(Y_i(u)-\alpha_u\right)\right)^\ell \leq C_\ell x^{\frac{\ell}{2}}\max_{1\leq i\leq x}\mathbb{E}\left|Y_i(u)-\alpha_u\right|^\ell,$$

where C_ℓ is a universal constant. Note that $\max_{u\in[0,1]}\mathbb{E}\left|Y_i(u)-\alpha_u\right|^\ell<\infty$, since $u\mapsto\alpha_u$ is continuous and bounded. Then, we deduce from the previous equalities that there exist two positive constants M_1 and M_2 such that

$$\delta_x Q_u V_p \leq \alpha_u^p V_p(x)+M_1 x^{p-1}+M_2.$$

To check the drift condition in **F1** for $m=1$, one can choose $\gamma>0$ such that $\lambda=\max_{u\in[0,1]}\alpha_u^p+\gamma<1$ and $b=M_2+\left(\frac{\gamma}{M_1}\right)^{p-1}$. In this case, the small set condition is satisfied for each finite set $\mathcal{C}$ with $\nu=\delta_0$ because

$$Q_u(x,0)\geq\left(1-\max_{u\in[0,1]}\alpha_u\right)^x\exp\left(-\max_{u\in[0,1]}\lambda_u\right)$$

for the INAR process and

$$Q_u(x,0)\geq\exp\left(-x\max_{u\in[0,1]}\alpha_u-\max_{u\in[0,1]}\lambda_u\right)$$

for the Poisson ARCH process. In both case, $\eta=\min_{u\in[0,1]}\min_{x\in\mathcal{C}}Q_u(x,0)>0$. This shows that assumption **F2** is satisfied by taking R large enough.
Finally, we show **F3**. Let $u,v\in[0,1]$. Denoting $\lambda_+=\max_{u\in[0,1]}\lambda_u$ and by μ_u the Poisson distribution of parameter λ_u, we have

$$\max_{u\in[0,1]}\mu_uV_p\leq 1+\mathbb{E}N_{\lambda_+}^p,\quad \|\mu_u-\mu_v\|_{V_p}\leq\sum_{k\geq 0}\frac{V_p(k)}{k!}\left(k\bar{\lambda}^{k-1}+\bar{\lambda}^k\right)\cdot|\lambda_u-\lambda_v|,$$

where $(N_t)_{t\geq 0}$ is Poisson process of intensity 1. Moreover, if ν_u denotes the Bernoulli distribution of parameter α_u, we have $\|\nu_u-\nu_v\|_{V_p}\leq 3|\alpha_u-\alpha_v|$. From Lemma 7 given below, we easily deduce that **F3** holds for $\widetilde{V}=CV_{p+1}$ where C is a positive real number. Note that we have $\sup_{u\in[0,1]}\pi_u\widetilde{V}<\infty$ because $\widetilde{V}$ also satisfies the drift and the small set conditions. Since all power functions satisfy the drift condition, the integrability condition required in the second point of Theorem 3 is automatic.

Lemma 7. *Let $X_1,X_2,\ldots,X_n,Y_1,Y_2,\ldots,Y_n$ be independent random variables such that $A_n=\max_{1\leq i\leq n}\mathbb{E}V(X_i)\vee\mathbb{E}V(Y_i)<\infty$ for $1\leq i\leq n$, with $V(x)=1+|x|^p$ and $p\geq 1$. Then we have*

$$\sup_{|f|\leq V}|\mathbb{E}f(X_1+\cdots+X_n)-\mathbb{E}f(Y_1+\cdots+Y_n)|\leq 2^{p-1}n^p\cdot A_n\cdot\max_{1\leq i\leq n}\|\mathbb{P}_{X_i}-\mathbb{P}_{Y_i}\|_V.$$

Proof of Lemma 7. Note first that if $|f(x)|\leq V(x)$ for all $x\in E$, then $|f(x+y)|\leq 2^{p-1}V(x)V(y)$. This leads to

$$\begin{aligned}
& |\mathbb{E}f(X_1+\cdots+X_n)-\mathbb{E}f(Y_1+\cdots+Y_n)| \\
\leq & \sum_{j=1}^n|\mathbb{E}f\left(X_1+\cdots+X_{j-1}+X_j+Y_{j+1}+\cdots+Y_n\right) \\
- & \mathbb{E}f\left(X_1+\cdots+X_{j-1}+Y_j+Y_{j+1}+\cdots+Y_n\right)| \\
\leq & 2^{p-1}\sum_{j=1}^n\|\mathbb{P}_{X_j}-\mathbb{P}_{Y_j}\|_V\cdot\mathbb{E}V\left(X_1+\cdots+X_{j-1}+Y_{j+1}+\cdots+Y_n\right) \\
\leq & 2^{p-1}(n-1)^{p-1}\sum_{j=1}^n\|\mathbb{P}_{X_j}-\mathbb{P}_{Y_j}\|_VA_n \\
\leq & 2^{p-1}n^pA_n\max_{1\leq i\leq n}\|\mathbb{P}_{X_i}-\mathbb{P}_{Y_i}\|_V.\square
\end{aligned}$$

17 Estimation of INAR/Poisson ARCH processes

Since we have

$$\mathbb{E}\left(X_{n,k}|X_{n,k-1}\right) = \alpha_{k/n}X_{n,k-1} + \lambda_{k/n},$$

a natural estimate is obtained by localized least squares. Set $a(u) = (\alpha_u, \lambda_u)'$ and $\mathcal{Y}_{n,i} = (1, X_{n,i-1})'$. Then we define

$$\begin{aligned}
\hat{a}(u) &= \arg\min_{\alpha} \sum_{i=2}^{n} e_i(u)\left(X_{n,i} - \mathcal{Y}_{n,i}'\alpha\right)^2 \\
&= \left(\sum_{i=2}^{n} e_i(u)\mathcal{Y}_{n,i}\mathcal{Y}_{n,i}'\right)^{-1} \sum_{i=2}^{n} e_i(u)X_{n,i}\mathcal{Y}_{n,i} := D_u^{-1}N_u.,
\end{aligned}$$

where the weights $e_i(u)$ are already defined in the main document.

The asymptotic behavior of this estimator is given in the following proposition.

Proposition 10. *Let $u \in (0,1)$. If $b \to 0$ and $nb \to \infty$, we have*

$$\lim_{n\to\infty} \sqrt{nb}\left(\hat{a}(u) - \mathbb{E}^{-1}(D_u)\mathbb{E}(N_u)\right) = \mathcal{N}_2\left(0, \Sigma(u)\right),$$

where $\mathcal{N}_2\left(0, \Sigma(u)\right)$ is the bivariate normal distribution with mean 0 and variance $\Sigma(u) = \int_{-1}^{1} K^2(v)dv\cdot$ $\Sigma_1(u)^{-1}\Sigma_2(u)\Sigma_1(u)^{-1}$,

$$\Sigma_1(u) = \mathbb{E}\left[\mathcal{Y}_1(u)\mathcal{Y}_1(u)'\right], \quad \Sigma_2(u) = \mathbb{E}\left[\left(X_1(u) - \mathcal{Y}_1(u)'a(u)\right)^2 \mathcal{Y}_1(u)\mathcal{Y}_1(u)'\right].$$

Moreover we have $\mathbb{E}^{-1}(D_u)\mathbb{E}(N_u) - a(u) = O(b)$.

We will apply Proposition 5 with $V(x) = 1 + |x|^p$ for some real number $p > 4$. But, first we will prove the second part of the proposition. We have from Theorem 4 and the compact support of the kernel K,

$$\sum_{i=2}^{n} e_i(u)\mathbb{E}\left[\mathcal{Y}_{n,i}\mathcal{Y}_{n,i}'\right] = \mathbb{E}\left[\mathcal{Y}_i(u)\mathcal{Y}_i(u)'\right] + O\left(b + \frac{1}{n}\right) = O\left(b\right).$$

In the same way, we have

$$\sum_{i=2}^{n} e_i(u)\mathbb{E}X_{n,i}\mathcal{Y}_{n,i} = \mathbb{E}X_i(u)\mathcal{Y}_i(u) + O\left(b + \frac{1}{n}\right) = O\left(b\right).$$

Moreover it is straightforward to prove that for all $n \geq 1$, $2 \leq k \leq n$ and $u \in [0,1]$, the matrices $\mathbb{E}\left[\mathcal{Y}_{n,k}\mathcal{Y}_{n,k}'\right]$ and $\mathbb{E}\left[\mathcal{Y}_k(u)\mathcal{Y}_k(u)'\right]$ are non degenerate. Using the approximations given above, this entails $\mathbb{E}^{-1}\left[D_u\right] = O(1)$ and also the second part of the proposition.

Next, using Proposition 5 and the martingale difference property, we have

$$\sqrt{nb}\sum_{i=2}^{n} e_i(u)\mathcal{Y}_{n,i}\left(X_{n,i} - \mathcal{Y}_{n,i}'a(i/n)\right) \to \mathcal{N}_2\left(0, \int_{-1}^{1} K^2(v)dv\cdot\Sigma_2(u)\right).$$

Using again Proposition 5, we have $D_u - \mathbb{E}(D_u) = o_{\mathbb{P}}(1)$. Then we deduce that

$$D_u^{-1}\sqrt{nb}\sum_{i=2}^n e_i(u)\mathcal{Y}_{n,i}\left(X_{n,i} - \mathcal{Y}_{n,i}'a(i/n)\right) \to \mathcal{N}_2\left(0, \Sigma(u)\right).$$

To end the proof, it remains to show that

$$D_u^{-1}\sum_{i=2}^n e_i(u)\mathcal{Y}_{n,i}\mathcal{Y}_{n,i}'a(i/n) - \mathbb{E}^{-1}(D_u)\mathbb{E}(N_u) = o_{\mathbb{P}}\left(\frac{1}{\sqrt{nb}}\right).$$

But the argument is similar to the proof of Theorem 5, point 3 (see assertion (4)). Hence we omit the details.☐

18 Local stationarity of SETAR models

We set $r\left(k/n, X_{n,k-1}\right) = X_{n,k} - \varepsilon_k$ and

$$Q_u(x, dy) = f_\varepsilon\left(y - r(u,x)\right)dy.$$

Note that $r(u,x) \leq \max\left(\max_{u\in[0,1]}|b(u)|, \max_{u\in[0,1]}|c(u)|\right) + \alpha|x|$. Assumption **F1** can be easily deduced with $m=1$ and $V(x) = 1+|x|^p$ or $V(x) = 1+|x|^{p+1}$. In the sequel, we set $V(x) = 1+|x|^p$. The levet sets for V are compact and for each compact set K, we have $\max_{u\in[0,1],x\in K}|r(u,x)| < \infty$. Morover $\eta = \inf_{x\in K,(y,u)\in[0,1]^2} f_\varepsilon(y - r(u,x)) > 0$ we have for all measurable sets A, $Q_u(x,A) \geq Q_u(x, A\cap[0,1]) \geq \eta\nu(A)$ where ν denotes the Lebesgue measure over $[0,1]$. This shows **F2** for any number $R > 0$. Next we check **F3**. From our assumptions, there exists $C > 0$ such that

$$\max_{u\in[0,1]}|r(u,x)| \leq C(1+|x|), \quad |r(u,x) - r(v,x)| \leq C(1+|x|)|u-v|.$$

Then we have, using the mean value theorem,

$$\begin{aligned}
\|\delta_x Q_u - \delta_y Q_v\|_V &\leq \int (1+|y|^p)|f_\varepsilon(y - r(u,x)) - f_\varepsilon(y - r(v,x))|dy \\
&\leq \int \left[1 + (|z| + C + C|x|)^p\right]\cdot\left|f_\varepsilon'(z)\right|dz\cdot\left|r(u,x) - r(v,x)\right|.
\end{aligned}$$

Then Assumption **F3** holds with $\widetilde{V}(x) = D\left(1+|x|^{p+1}\right)$ for a suitable chosen real number $D > 0$. One can also check the condition of the second point in Theorem 3. Indeed setting $L(x) = K(1+|x|)$ for K large enough, the arguments used previously give

$$\|\delta_x Q_u - \delta_x Q_v\|_{TV} \leq L(x)|u-v|.$$

Moreover, from the drift condition, we have $\sup_{u\in[0,1]}\pi_u\widetilde{V} < \infty$ and the finiteness of

$$\sup_{\substack{u\in[0,1]\\ 1\leq\ell'\leq\ell}} \mathbb{E}\left[L\left(X_\ell(u)\right)V\left(X_{\ell'}(u)\right)\right]$$

is automatic.

19 Proof of Proposition 4

The proof is based on the central limit theorem for strongly mixing triangular arrays given in Rio (1995), Corollary 1. Note that, the strong mixing coefficients are upper bounds of the $\beta-$mixing coefficients. Moreover, it is clear that the triangular array $\{Z_{n,i} : j \leq i \leq n, n \geq j\}$ is still $\beta-$mixing with a geometric rate of convergence.

1. Suppose first that $\sigma^2 > 0$. We set $H_{n,i} = \frac{1}{\sqrt{n}\sigma}\left[Z_{n,i} - \mathbb{E}Z_{n,i}\right]$ and we study the asymptotic behavior of $\sum_{i=1}^n H_{n,i}$ which is the same than $\sum_{i=j}^n H_{n,i}$ (by convention, we assume that $(X_{n,i})_{i\leq 0}$ is a path of the stationary Markov chain with transition Q_0). For $1 \leq i \leq n$, we also set $V_{n,i} = \mathrm{Var}\left(\sum_{k=1}^i H_{n,i}\right)$,

$$Q_{n,i}(x) = \sup\left\{t \in \mathbb{R}_+ : \mathbb{P}\left(|H_{n,i}| > t\right) > x\right\}, \quad x \in (0,1)$$

and we denote by β_n^{-1} the inverse function of the $\beta-$mixing rate function (i.e $x \mapsto \beta_n([x])$). One can check that there exist $C > 0$ such that for all $(n,v) \in \mathbb{N}^* \times (0,1)$ and , $\beta_n^{-1}(v) \leq C(-\log(v) + 1)$. Moreover, from our assumptions, there also exists $\rho \in (0,1)$ and $C' > 0$ such that $\beta_n(j) \leq C'\rho^j$, if $j \leq n$. To apply the result of Rio (1995), we need to check the two following conditions.

$$\limsup_{n\to\infty} \max_{1\leq i\leq n} \frac{V_{n,i}}{V_{n,n}} < \infty, \tag{12}$$

$$\lim_{n\to\infty} V_{n,n}^{-3/2} \sum_{i=1}^n \int_0^1 \beta_n^{-1}(x) Q_{n,i}^2(x) \inf\left(\beta_n^{-1}(x) Q_{n,i}(x), \sqrt{V_{n,n}}\right) dx = 0. \tag{13}$$

- First, we check that $\sup_{n\geq 1} \max_{1\leq i\leq n} V_{n,i} < \infty$. It is enough to bound the covariances $\mathrm{Cov}\left(Z_{n,i}, Z_{n,i+k}\right)$. Using the covariance inequality for strong mixing coefficients (see Doukhan (1994), Theorem 3), we get

$$\left|\mathrm{Cov}\left(Z_{n,i}, Z_{n,i+k}\right)\right| \leq 8\beta_n(k) \max_{1\leq i\leq n} \|Z_{n,i}\|_{2+\delta}^2.$$

  Then the result follows from the assumption made on V, which ensures that $\max_{1\leq i\leq n} \|Z_{n,i}\|_{2+\delta} = O(1)$ and the geometric decay of the $\beta-$mixing coefficients.

- Next, we check that $\lim_{n\to\infty} V_{n,n} = 1$. From the mixing properties given in Proposition 3 of the paper and the covariance inequality for strong mixing sequences (see the previous point), we have

$$\sup_{u\in[0,1]} \sum_{k\in\mathbb{Z}} \left|\mathrm{Cov}\left(Z_0(u), Z_k(u)\right)\right| < \infty.$$

  Now let λ be a positive real number. We first fix a positive integer K such that

$$\sum_{|k|>K} \int_0^1 \left|\mathrm{Cov}\left(Z_0(u), Z_k(u)\right)\right| du + \max_{1\leq i\leq n} \sum_{\substack{1\leq j\leq n \\ |j-i|>K}} \left|\mathrm{Cov}\left(Z_{n,i}, Z_{n,j}\right)\right| < \lambda/2.$$

Now there exists a positive integer n_0 such that for all $n \geq n_0$,

$$\left| \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{1 \leq j \leq n \\ |j-i| \leq K}} \mathrm{Cov}\left(Z_0(i/n), Z_{j-i}(i/n)\right) - \int_0^1 \sum_{j=-K}^{K} \mathrm{Cov}\left(Z_0(u), Z_j(u)\right) du \right| \leq \lambda/4 \quad (14)$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{1 \leq j \leq n| \\ |j-i| \leq K}} \left|\mathrm{Cov}\left(Z_{n,i}, Z_{n,j}\right) - \mathrm{Cov}\left(Z_0(i/n), Z_{j-i}(i/n)\right)\right| \leq \lambda/4. \quad (15)$$

Indeed, (14) can be proved using the continuity of $u \mapsto \mathrm{Cov}\left(Z_0(u), Z_j(u)\right)$ which follows from the local stationarity property, the continuity of f with respect to its first argument and the Lebesgue theorem. Moreover, (15) can be proved using the approximation with stationary Markov chains, this approximation being of the order $1/n$. Finally, we conclude that for $n \geq n_0$, $|V_{n,n} - 1| \leq \lambda$. This justifies the limiting behavior of $V_{n,n}$. Moreover, using the previous point, assertion (12) follows.

- Finally, we prove (13). Using the Markov inequality, there exists a constant $C > 0$ such that

$$\mathbb{P}\left(|H_{n,i}| > y\right) \leq \frac{C}{n^{1+\delta/2} y^{2+\delta}}.$$

Then we have $\max_{1 \leq i \leq n} Q_{n,i}(x) \leq \widetilde{C} x^{-\frac{1}{2+\delta}} n^{-\frac{1}{2}}$ and assertion (13) follows from the Lebesgue theorem.

The proof of point 1 is complete when $\sigma > 0$. If $\sigma = 0$, it is easily seen from the previous arguments that

$$\lim_{n \to \infty} \frac{1}{n} \mathrm{Var}\left(\sum_{i=1}^{n} Z_{n,i}\right) = \sigma^2 = 0$$

and the limiting distribution is the Dirac mass at point 0.

2. The proof for the second point is similar to the first one by setting

$$H_{n,k} = \frac{1}{\sqrt{nb}\,\sigma(u) \|K\|_{\mathbb{L}^2}} K\left(\frac{u - k/n}{b}\right) \left[Z_{n,k} - \mathbb{E} Z_{n,k}\right],$$

where $\|K\|_{\mathbb{L}^2} = \sqrt{\int_{-1}^{1} K(v)^2 dv}$. Using the same notations as for point 1 and the compact support of the kernel, one can show that $\lim_{n \to \infty} V_{n,n} = 1$ and condition (12) is satisfied. Moreover,

$$Q_{n,i}(x) \leq C x^{-\frac{1}{2+\delta}} (nb)^{-\frac{1}{2}} \mathbb{1}_{|u-i/n| \leq b}$$

and the proof of (13) is similar to the proof of point 1. Details are omitted.□

20 Localized Maximum likelihood estimator

20.1 Asymptotic properties

Proof of Theorem 4 The proof follows that of Dahlhaus et al. (2017), Theorem 5.2 and 5.4. We check the main arguments. We set $\mathcal{L}(\theta) = \mathbb{E}\left[S\left(\theta, X_0(u), X_1(u)\right)\right]$.

1. From Theorem 4 and Proposition 4, we have for each $\theta \in \Theta$,

$$\mathcal{L}_n(\theta) - \mathcal{L}(\theta) = O_{\mathbb{P}}\left(b + \frac{1}{\sqrt{nb}}\right).$$

Moreover, using our assumptions and the Lebesgue theorem, the function $\theta \mapsto \mathcal{L}(\theta)$ is continuous. Next, we study the stochastic equicontinuity of $\mathcal{L}_n$. Let $\epsilon, \delta > 0$. We have,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{|\theta-\theta'|\leq\delta}\left|\mathcal{L}_n(\theta) - \mathcal{L}_n(\theta')\right| > \epsilon\right) \\
\leq\ & \frac{1}{\epsilon n}\sum_{k=2}^{n} K_b(u-k/n)\mathbb{E}\left[\sup_{|\theta-\theta'|\leq\delta}\left|S\left(\theta, X_{n,k-1}, X_{n,k}\right) - S\left(\theta', X_{n,k-1}, X_{n,k}\right)\right|\right] \\
=\ & \frac{1}{\epsilon n}\sum_{k=2}^{n} K_b(u-k/n)\mathbb{E}\left[\sup_{|\theta-\theta'|\leq\delta}\left|S\left(\theta, X_{k-1}(u), X_k(u)\right) - S\left(\theta', X_{k-1}(u), X_k(u)\right)\right|\right] + O(b).
\end{aligned}$$

From our assumptions and the Lebesgue theorem, we deduce that

$$\lim_{\delta\to 0}\lim\sup_{n\to\infty}\mathbb{P}\left(\sup_{|\theta-\theta'|\leq\delta}\left|\mathcal{L}_n(\theta) - \mathcal{L}_n(\theta')\right| > \epsilon\right) = 0.$$

We deduce that

$$\max_{\theta\in\Theta}\left|\mathcal{L}_n(\theta) - \mathcal{L}(\theta)\right| = O_{\mathbb{P}}(1).$$

By standard arguments, we get the consistency of the estimator $\hat{\theta}(u)$.

2. From the assumptions and the Lebesgue theorem, the function $\mathcal{L}$ is two times continuously differentiable on Θ and $\nabla^2\mathcal{L}(\theta) = \mathbb{E}\left[\nabla_1^2 S\left(\theta, X_0(u), X_1(u)\right)\right]$. Using the same arguments as in the previous point, one can show that

$$\max_{\theta\in\Theta}\left|\nabla^2\mathcal{L}_n(\theta) - \nabla^2\mathcal{L}(\theta)\right| = o_{\mathbb{P}}(1).$$

Next, using Proposition 5 given in the paper, and the martingale difference property for the stationary approximation, we have

$$\sqrt{nb}\left[\nabla\mathcal{L}_n(\theta_0(u)) - \mathbb{E}\mathcal{L}_n(\theta_0(u))\right] \Rightarrow \mathcal{N}\left(0, \int K^2(x)dx I(u)\right),$$

with

$$I(u) = \mathbb{E}\left[\nabla_1 S\left(\theta_0(u), X_0(u), X_1(u)\right)\nabla_1 S\left(\theta_0(u), X_0(u), X_1(u)\right)'\right].$$

Note that $I(u)$ is also equal to the quantity $M(u)$ defined in the statement of the theorem. Finally, we derive an expansion for the bias. Using our approximation results and the symmetry of the kernel, we have

$$\begin{aligned}
& \mathbb{E}\nabla\mathcal{L}_n(\theta_0(u)) - g_u(u) \\
= \quad & \frac{1}{n}\sum_{j=2}^{n} K_b(u-j/n)(j/n-u)g_u'(u) + \frac{1}{2n}\sum_{j=2}^{n} K_b(u-j/n)(u-j/n)^2 g_u''(u) + o(b^2) + O(1/n) \\
= \quad & \frac{1}{2}b^2 g_u''(u)\int v^2 K(v)dv + o(b^2) + O(1/n).
\end{aligned}$$

The scheme of the rest of proof is similar to that of Dahlhaus et al. (2017).□

20.2 Examples

Binary time series For the binary time series defined by the equation (4) in the paper, we consider

$$Q_u(x,y) = F\left(\theta_{00}(u) + \sum_{j=1}^{p}\theta_{0j}(u)x_{p+1-j}\right)\prod_{i=1}^{p-1}\mathbb{1}_{y_i=x_{i+1}}, \quad x,y\in\{0,1\}^p \text{ s.t. } y_p = 1.$$

It is easily seen that Q_u^p has positive entries. If $u\mapsto\theta_0(u)$ is two times continuously differentiable, then assumptions **A1-A2** of the paper are satisfied. We remind that assumptions **F1-F3** are then satisfied, setting $V,\widetilde{V}$ and L to 1. Assumptions **L1**(2) and **L2-L4** are then valid. To check **L5**, one can use the fact that for a finite-state irreducible Markov chain, the invariant probability is a $\mathcal{C}^\infty-$function of the transition matrix. See Cao (1998) for details. A general result is also given in Truquet (2017b) for general state spaces.

Poisson ARCH process We consider the case of one lag for simplicity but extension to several lags is possible. We have

$$S(\theta,x,y) = -\theta_0 - \theta_1 x + y\log(\theta_0+\theta_1 x) - \log(y!).$$

Assumption **L1**(2) holds true. We have seen that the model is locally stationary in V norms for each function $V(x) = 1 + x^p$. The assumption **L2** and **L4** are satisfied if p is large enough ($p\geq 4$ is sufficient). Assumption **L3** is satisfied if $\theta_{00}(u)$ and $\theta_{01}(u)$ are positive. Remember that $\max_{u\in[0,1]}\theta_{01}(u) < 1$. Identification of the parameter is similar to the stationary case. For Assumption **L5**, we assume that $u\mapsto\theta_0(u)$ is two times continuously differentiable and we use a result given in Truquet (2017b). See Proposition 4 and Section 4.3 of this reference. This result guarantees that $u\mapsto\int f d\pi_{u,2}$ is two times continuously differentiable for any function f such that $|f(x,y)|\leq C(1+x^p+y^p)$, whatever the value of the positive integer p.

21 Proof of Proposition 5

The key argument for the proof is to show that $\lim_{n\to\infty}\hat{Q}_u = Q_u$ a.s. From this convergence, we will deduce that for almost ω, there exists an integer n_0 such that for $n\geq n_0$, a Markov chain with transition $\hat{Q}_{u,\omega}$ is geometrically $\phi-$mixing. Taking in account of the results of Theorem 2, this almost sure convergence can be obtained if we show that for any function $f: E^2 \rightarrow \mathbb{R}$, we have

$$A_n := \frac{1}{n}\sum_{j=2}^{n} K_b\left(u - j/n\right)\left[f\left(X_{n,j-1}, X_{n,j}\right) - \mathbb{E}f\left(X_{n,j-1}, X_{n,j}\right)\right] \rightarrow_{n\rightarrow\infty} 0. \text{ a.s.}$$

Using the Borel-Canteli Lemma, it is sufficient to show that for any $\delta > 0$, we have $\sum_{n\geq 1}\mathbb{P}(A_n > \delta) < \infty$. But, remembering that $nb^{1+\epsilon} \rightarrow \infty$, this assertion can be shown using the exponential inequality given in (1) with a choice $q \sim n^{\alpha}$, $\lambda = \delta nb$ and $0 < \alpha < \frac{\epsilon}{1+\epsilon}$.

Now, conditioning with respect to a path of the triangular array, one can consider that $\hat{Q}_u$ is deterministic and convergent towards Q_u. Remind that there exists a positive integer m such that Q_u^m is contracting in total variation. Now we have $\mathbb{P}^*\left(X_i^* = x\right) = \hat{\pi}_u\hat{Q}_u^i$ and using a contraction argument already used in the proof of Theorem 1 of the paper, one can show that

$$\|\pi_u - \hat{\pi}_u\hat{Q}_u^{s+mk}\|_{TV} \leq \frac{1}{1-c}\max_{x\in E}\|\hat{Q}_u^m(x,\cdot) - Q_u^m(x,\cdot)\|_{TV} + c^k,$$

with $c := c(Q_u^m) < 1$ denotes the Dobrushin's contraction coefficient of Q_u^m. Then we deduce that there exists a constant $D > 0$ such that

$$\frac{1}{n}\sum_{j=2}^{n} K_b\left(u - j/n\right)\left|\mathbb{P}^*\left(X_i^* = x\right) - \pi_u(x)\right| \leq D\left(\frac{1}{nb} + \max_{x\in E}\|\hat{Q}_u^m(x,\cdot) - Q_u^m(x,\cdot)\|_{TV}\right) = o(1).$$

On the other hand, if n is large enough, a Markov chain with transition $\hat{Q}_u$ will be also geometrically $\phi-$mixing because $\hat{Q}_u^m$ will be contracting. Using covariance inequalities, we obtain

$$\frac{1}{n}\sum_{j=2}^{n} K_b\left(u - j/n\right)\left(\mathbb{1}_{\{X_j^*=x\}} - \mathbb{P}^*\left(X_i^* = x\right)\right) = o_{\mathbb{P}^*}(1).$$

Then we deduce that $\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{j=2}^{n} K_b\left(u - j/n\right)\mathbb{1}_{\{X_j^*=x\}} = \pi_u(x)$ in $\mathbb{P}^*$ probability. Finally, using the decomposition

$$\sqrt{nb}\left(\hat{Q}_u^*(x,y) - \hat{Q}_u(x,y)\right) = \frac{\frac{1}{\sqrt{nb}}\sum_{j=2}^{n} K\left(\frac{u-j/n}{b}\right)\mathbb{1}_{\{X_{j-1}^*=x\}}\left[\mathbb{1}_{\{X_j^*=y\}} - \hat{Q}_u(x,y)\right]}{\frac{1}{n}\sum_{j=2}^{n} K_b\left(u - j/n\right)\mathbb{1}_{\{X_j^*=x\}}}$$

and the central limit theorem for triangular arrays of martingales, we deduce the result because the numerator is asymptotically Gaussian with mean 0 and variance

$$\lim_{n\rightarrow\infty}\frac{1}{nb}\sum_{j=2}^{n} K^2\left(\frac{u-j/n}{b}\right)\mathbb{1}_{\{X_{j-1}^*=x\}}\left[\hat{Q}_u(x,y) - \hat{Q}_u(x,y)^2\right] = \int_{-1}^{1} K^2(v)dv\cdot\left[Q_u(x,y) - Q_u(x,y)^2\right].$$

The previous limit holds in $\mathbb{P}^*-$probability. Collecting all the previous points, the proof of the convergence of the bootstrap estimator is now complete. The second part of the proposition will follow from the asymptotic normality of $\hat{Q}_u$ given in Theorem 5 but it is necessary to show that the bias is of order $o(b)$. Using a Taylor expansion, we have

$$\frac{\mathbb{E}\hat{\pi}_{u,2}(x,y)}{\mathbb{E}\hat{\pi}_u(x)} = \frac{\sum_{i=1}^{n-1} e_i(u)Q_{i/n}(x,y)\pi_{i-1}^{(n)}(x)}{\sum_{i=1}^{n-1} e_i(u)\pi_{i-1}^{(n)}(x)} = Q_u(x,y) + o(b).\square$$

22 An additional example for Theorem 4: the random walk on the positive integers

Let $p, q, r : [0,1] \to (0,1)$ three $\kappa-$Hölder continuous functions such that $p(u)+q(u)+r(u) = 1$ and $\frac{p(u)}{q(u)} < 1$. For $x \in \mathbb{N}^*$, we set $Q_u(x,x) = r(u)$, $Q_u(x,x+1) = p(u)$ and $Q_u(x,x-1) = q(u)$. Finally $Q_u(0,1) = 1 - Q_u(0,0) = p(u)$. In the homogeneous case, geometric ergodicity holds under the condition $p < q$. See Meyn and Tweedie (2009), Chapter 15. In this case the function V defined by $V(x) = z^x$ is a Foster-Lyapunov function if $1 < z < q/p$. For the non-homogeneous case, let $z \in (1,e)$ where $e = \min_{u\in[0,1]} q(u)/p(u)$. We set $\gamma = \max_{u\in[0,1]} \left\{r(u)+p(u)z+q(u)z^{-1}\right\}$ and $\overline{p} = \max_{u\in[0,1]} p(u)$. Note that

$$\gamma \leq 1 + \overline{p}(z-1)\max_{u\in[0,1]}\left[1 - \frac{q(u)}{p(u)z}\right] \leq 1 + \overline{p}(z-1)\left[1 - \frac{e}{z}\right] < 1.$$

Then we have $Q_uV(x) \leq \gamma V(x)$ for all $x > 0$ and $Q_uV(0) = p(u)z + (1-p(u)) \leq c = \overline{p}(z-1)+1$. For an integer $m \geq 1$, we have $Q_{u_1}\cdots Q_{u_m}V \leq \gamma^m V + \frac{c}{1-\gamma}$. If m is large enough, we have $\frac{2c}{(1-\gamma)(1-\gamma^m)V(m)} < 1$. Moreover, for such m, if $R = V(m)$, we have $\{V \leq R\} = \{0,1,\ldots,m\}$ and if $x = 0,\ldots,m$, we have $\delta_x Q_{u_1}\cdots Q_{u_m} \geq \eta\delta_0$ for a $\eta > 0$. Assumption **F3** is immediate. Moreover the additional condition in the second point of Theorem 4 is automatically checked with a constant function L.
However this example is more illustrative. Indeed parameters $p(u)$ and $q(u)$ can be directly estimated by

$$\hat{p}(u) = \sum_{i=1}^{n-1} e_i(u)\mathbb{1}_{X_{n,i+1}-X_{n,i}=1}, \quad \hat{q}(u) = \sum_{i=1}^{n-1} e_i(u)\mathbb{1}_{X_{n,i+1}-X_{n,i}=-1},$$

where the weights $e_i(u)$ are defined as in Theorem 2. The indicators are independent Bernoulli random variables with parameter $p\left(\frac{i+1}{n}\right)$ or $q\left(\frac{i+1}{n}\right)$ and the asymptotic behavior of the estimates is straightforward.

23 An additional real data set example for finite-state Markov chains

We illustrate our methods with an analysis of daily rainfall data recorded in London's Saint-James Park station, between January 2017 and September 2017. Data are available from `www.ogimet.com/indicativos.phtml.en`. The sample size is $n = 270$ and we build a binary time series by setting $X_i = 1$ if rainfall has been recorded at day i and $X_i = 0$ otherwise. The autocorrelogram suggests that the first autocorrelations are significant. We then fit a time-inhomogeneous Markov chain of order 1. The selected bandwidth is $\hat{b} = 0.41$. The autocorrelogram of the residuals does not suggest a remaining dependence structure. The estimated values of the transition probabilities $u \mapsto Q_u(0,0)$ and $u \mapsto Q_u(1,1)$ are represented in Figure 1. It seems that the transition probabilities change smoothly with time. A notable particularity is the maximum/minimum values of the diagonal elements of the stochastic matrix around the observation $i = 150$. Larger probabilities (lower probabilities resp.) of getting another dry day (rainy day resp.) coincides with the spring/summer period. This seasonal behavior is of course expected and is recovered by the model which seems to be a good candidate for extracting such features from the data.

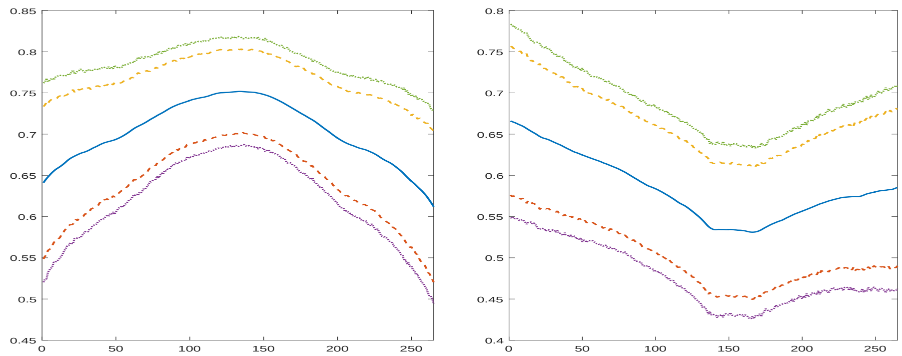

Figure 1: Estimation of $u \mapsto Q_u(0,0)$ (left) and $u \mapsto Q_u(1,1)$ (right). The estimates are given by the full line and the dashed lines (dotted lines resp.) represent the bootstrap pointwise confidence intervals at level 80% (90% resp.) and which are estimated using $B = 5000$ bootstrap samples.

References

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

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

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

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

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

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

C. Francq and J-M. Zakoïan. *GARCH models: structure, statistical inference and financial applications*. Wiley, 2010.

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

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

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. Birkhäuser,/Springer Basel AG, Basel, 2011.

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

E. Rio. About the lindeberg method for strongly mixing sequences. *ESAIM, Probability and Statistics*, 1:35–61, 1995.

E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*. Springer, 1999.

E. Rio. Inequalities and limit theorems for weakly dependent sequences. https://cel.archives-ouvertes/cel-00867106, 2013.

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

L. Truquet. Parameter stability and semiparametric inference in time-varying arch models. *J. R. Stat. Soc. Ser. B*, 79:1391–1414, 2017a.

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

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

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

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