Iterations of dependent random maps and exogeneity in nonlinear dynamics

Max Zinsou Debaly *   Lionel Truquet †

Abstract

We discuss existence and uniqueness of stationary and ergodic nonlinear autoregressive processes when exogenous regressors are incorporated in the dynamic. To this end, we consider the convergence of the backward iterations of dependent random maps. In particular, we give a new result when the classical condition of contraction on average is replaced with a contraction in conditional expectation. Under some conditions, we also derive an explicit control of the functional dependence of Wu (2005) which guarantees a wide range of statistical applications. Our results are illustrated with CHARME models, GARCH processes, count time series, binary choice models and categorical time series for which we provide many extensions of existing results.

2010 Mathematics Subject Classification: Primary 62M10; secondary 60G10, 60B12.
Keywords and Phrases: time series, random maps, ergodicity, dependence.

1 Introduction

Among the various contributions devoted to time series analysis, theoretical results justifying stationary and ergodicity properties of some standard stochastic processes when exogenous covariates are incorporated in the dynamic are rather scarce. A notable exception concerns linear models, such as VARMA processes, for which such properties are a consequence of the linearity. See for instance Lütkepohl (2005), a standard reference for multivariate time series models. Moreover, linear models represent a very simple setup for discussing various exogeneity notions found in the literature. See for instance Engle et al. (1983). For non linear dynamics, a few contributions consider the problem of exogenous regressors. For general GARCH type processes, Francq and Thieu (2019) recently studied stationary conditions when the noise and the covariate process form a stationary process. Agosto et al. (2016) considered a Poisson auotregressive process with exogenous regressors (PARX models), under a Markov chain assumption for the covariate process. de Jong and Woutersen (2011) consider the case of dynamic binary choice models and provide results about stationarity and mixing properties of a $0/1-$valued time series which is autoregressive and defined conditionally to some exogenous regressors. Fokianos and Truquet (2018) studied stationarity and

*CREST-ENSAI, UMR CNRS 9194, Campus de Ker-Lann, rue Blaise Pascal, BP 37203, 35172 Bruz cedex, France.
†CREST-ENSAI, UMR CNRS 9194, Campus de Ker-Lann, rue Blaise Pascal, BP 37203, 35172 Bruz cedex, France. *Email: lionel.truquet@ensai.fr*

ergodicity of general categorical time series defined conditionally to a strictly exogenous covariate process.

In this paper, we give general results for getting stationarity, ergodicity and dependence properties for general non linear dynamics defined in term of iterations of random maps. For simplicity, we explain our setup with the following example which represents the basis for studying other processes. Let us consider the following dynamic

$$X_t = F\left(X_{t-1}, Z_{t-1}, \varepsilon_t\right), \quad t \in \mathbb{Z}, \tag{1}$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a covariate process and $(\varepsilon_t)_{t\in\mathbb{Z}}$ a noise process. One can note that $X_t = f_t(X_{t-1})$ for the random function defined by $f_t(x) = F\left(x, Z_{t-1}, \varepsilon_t\right)$. The sequence $(f_t)_{t\in\mathbb{Z}}$ is a sequence of dependent random maps even if the ε_t's are i.i.d. because the Z_t's exhibit temporal dependence. A key point for getting existence of a stationary solution in (1) is to control the behavior of the backward iterations $\{f_t \circ f_{t-1} \circ \cdots \circ f_{t-n}(x) : n \geq 1\}$. The convergence of such iterations of random maps have been extensively studied in the independent case. In this case, the process $(X_t)_{t\in\mathbb{Z}}$ is a Markov chain. We defer the reader to Letac (1986) and Diaconis and Freedman (1999) for seminal papers on iterated independent random maps and to Wu and Shao (2004) for additional results useful in a time series context. The last contribution is particularly interesting for getting existence of some moments for the marginal X_t and also some dependence properties for the process $(X_t)_{t\in\mathbb{Z}}$ that are often needed for statistical applications. All these contributions use average contraction conditions and the interested reader is deferred to the interesting survey of Stenflo (2012) for an overview of the available results. There also exist some contributions studying the more general case of iterated stationary random maps $(f_t)_{t\in\mathbb{Z}}$. For instance, Borovkov (1998) gives many results for studying what he calls stochastically recursive sequences, when the independence assumption is removed. See also Iosifescu (2009) for a survey of some available results. The results obtained in the dependent case are based on Lyapunov type exponents and the convergence of the backward iterations is only studied almost surely. Let us remind the following result which can be found in Elton (1990) (see also Iosifescu (2009), Theorem 6.2) and which generalizes a widely known result given in Brandt (1986) or Bougerol and Picard (1992) for iterations of affine random maps.

We first introduce some notations. We assume that $f_t : E \to E$ are random Lipschitz functions where E denotes a locally compact Polish space endowed with a metric d. We define the Lipschitz constant of a measurable function $g : E \to E$ by

$$c(g) := \sup_{x \neq y \in E} \frac{d\left(g(x), g(y)\right)}{d(x, y)}.$$

Moreover, for any integers $s < t$, we set $f_s^t = f_t \circ \cdots \circ f_s$.

Theorem 1. *Assume that the process $\left((Z_t, \varepsilon_t)\right)_{t\in\mathbb{Z}}$ in (1) is stationary and ergodic. Assume further that $\mathbb{E}\left[\log^+ c(f_0)\right] < \infty$ and $\mathbb{E}\left[\log^+ d(y_0, f_0(y_0))\right] < \infty$ for some point $y_0 \in E$.*

1. *There exists a constant $\chi \in \mathbb{R} \cup \{-\infty\}$ called Lyapunov exponent and such that*

$$\lim_{n\to\infty} \frac{1}{n} \log c\left(f_1^n\right) = \chi \ a.s.$$

Moreover

$$\gamma = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left[\log c\left(f_1^n\right)\right] = \inf_{n\geq 1} \frac{1}{n} \mathbb{E}\left[\log c\left(f_1^n\right)\right].$$

2. *If the constant χ is negative, then the almost sure limit $f_{-\infty}^t = \lim_{k\to\infty} f_{t-k}^t(x)$ exists for any $x \in E$ and does not depend on x. Setting $X_t = f_{-\infty}^t$, the process $(X_t)_{t\in\mathbb{Z}}$ is stationary and ergodic and satisfies the recursions (1). Moreover, $(X_t)_{t\in\mathbb{Z}}$ is the unique stationary process satisfying (1).*

The affine random maps version of this result has been applied recently by Francq and Thieu (2019) for studying stationarity of asymmetric power GARCH processes. For non linear random maps, Theorem 1 is less known in the time series literature. To illustrate its interest, we will see in Section 4.4 how Theorem 1 can be applied to binary time series and lead to an improvement of a result of de Jong and Woutersen (2011). However the result presented above has several limitations.

1. First, it requires the random maps f_t to be Lipschitz. Such property is not always valid, for instance for some discrete-valued time series models considered in this paper. When there is no exogenous covariates, Davis and Liu (2016) studied integer-valued time series by using a different contraction result, developed by Wu and Shao (2004).

2. Moreover, existence of some moments for the marginal distributions that are sometimes necessary for statistical applications cannot be obtained directly from this result.

3. Finally, especially for autoregressions with several lags, it is not straightforward to get an explicit condition on the parameters of the model to ensure that $\chi < 0$.

To overcome these drawbacks, we will adapt the approach used by Wu and Shao (2004) for independent random maps to the case of dependent random maps. Our main result is obtained by replacing the usual contraction on average condition by a contraction in conditional expectation. The assumptions we use are very simple to check and the proof of our main result is straightforward but its merit is to provide an elegant way for presenting a general approach which encompasses most of the previous attempts to include exogenous regressors in nonlinear dynamics. Moreover, we will discuss how to control the functional dependence measure of $(X_t)_{t\in\mathbb{Z}}$ introduced by Wu (2005), a dependence notion which is an alternative to the standard strong mixing condition and which can be more easily checked for iterations of contracting random maps. Let us mention that even in the independent case, mixing properties of the process $(X_t)_{t\in\mathbb{Z}}$ require restrictive assumptions on the noise distribution otherwise such properties may fail. We defer the reader to the standard textbook of Doukhan (1994), section 2.4 for mixing properties of iterations of independent random functions. Moreover, in the dependent case, as in (1), getting usual strong mixing properties seems to be harder because the process $(X_t)_{t\in\mathbb{Z}}$ does not have a Markov structure in general.

This paper is mainly motivated by dynamics of type (1) with covariates that are exogenous in a weak sense, assuming that at any time t, the noise ε_t is independent from the past information $\sigma\left((Z_s, \varepsilon_s) : s \leq t-1\right)$. This independence assumption is substantially weaker than the independence between the two processes ε and Z. The latter independence condition implies strict exogeneity, a notion initially defined by Sims (1972) and extended to general models by Chamberlain (1982). Strict exogeneity is useful for deriving the conditional likelihood of the X_t's conditionally to the Z_t's. However, strict exogeneity is a rather strong assumption. Under additional regularity conditions on the model, Chamberlain (1982) has shown that this assumption is equivalent to the non Granger-causality, i.e. Z_t is independent of $(X_s)_{s\leq t}$ conditionally to $(Z_s)_{s\leq t-1}$. It roughly means that the covariate process $(Z_t)_{t\in\mathbb{Z}}$ evolves in a totally autonomous way. In contrast, our exogeneity condition allows general covariates of the form $Z_t = H\left(\eta_t, \eta_{t-1}, \ldots\right)$ with H a measurable function and a sequence $((\eta_t, \varepsilon_t))_{t\in\mathbb{Z}}$ of i.i.d. random vectors, ε_t being possibly correlated with η_t. The error ε_t can then still have an influence future values of the covariates. For linear models, the two technical independence conditions discussed above between the noise and the covariate processes are often used as a distinction between weak and strict exogeneity. See for instance Lütkepohl (2005), Section 10.2. Let us mention that there exist additional concepts of exogeneity that are introduced and discussed in Engle et al. (1983) in particular another notion of weak exogeneity. However this notion is related to the estimation of a specific parameter of the conditional distribution for the bivariate process (X_t, Z_t) and it is necessary to specify the joint dynamic of the process. Since we do not want to consider specific dynamics for the covariate process, we will not use it in this paper. Inclusion of weakly exogenous regressors motivates our approach which is based on conditional average contraction conditions. But our results can be also applied without referring to these concepts of exogeneity, i.e. when $((Z_{t-1}, \varepsilon_t))_{t\in\mathbb{Z}}$ is a general stationary and ergodic process in (1). However, in the latter case, a closed form expression for the conditional distribution of X_t given $\mathcal{F}_{t-1}$ cannot be obtained directly from the recursions (1).

The paper is organized as follows. In Section 2, we give our general result for defining stationary and ergodic solutions for recursions of type (1). In Section 3, we study the stochastic dependence of the process using the functional dependence measure of Wu (2005). This dependence notion has many advantages with respect to the standard strong mixing condition, the latter being difficult to check or invalid for random iterations. Many examples are given in Section 4, we revisit many nonlinear dynamics discussed recently in the literature but also new ones. Our contribution is then the first one presenting a unified framework for inclusion of covariates in nonlinear dynamics. Note we do not consider particular statistical applications, any of them which require either existence of ergodic paths and/or a control of the functional dependence measure can be considered, some references are listed in Section 3.

2 General result

Let $(f_t)_{t\in\mathbb{Z}}$ a sequence of random maps defined on a Polish space (E, d) and taking values in the same space. We assume for convenience that $f_t = F(\cdot, \zeta_t)$ where $(\zeta_t)_{t\in\mathbb{Z}}$ is a stationary process taking values in a measurable space E' and $G : E \times E' \to E$ is a measurable application. In connection with our initial example (1), we have $\zeta_t = (Z_{t-1}, \varepsilon_t)$. For $s < t$, we set $f_s^t = f_t \circ f_s^{t-1}$ with the convention $f_t^t = f_t$ and $f_t^{t-1}(x) = x$. Moreover, we consider a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ for which $(\zeta_t)_{t\in\mathbb{Z}}$ is adapted.

We assume that there exist $p \geq 1$, $L \geq 1$, $\kappa \in (0, 1)$, an integer $m \geq 1$ and $y_0 \in E$ such that the following conditions will be fulfilled.

A1 For any $x \in E$, we have $\mathbb{E}\left[d\left(f_0(x), y_0\right)\right] < \infty$.

A2 We have for all $t \in \mathbb{Z}$ and almost surely,

$$\mathbb{E}\left[d^p\left(f_t(x), f_t(y)\right)|\mathcal{F}_{t-1}\right] \leq L^p d^p(x, y) \text{ and } \mathbb{E}\left[d^p\left(f_t^{t+m-1}(x), f_t^{t+m-1}(y)\right)|\mathcal{F}_{t-1}\right] \leq \kappa^p d^p(x, y).$$

We denote by $\|X\|_p$ the $\mathbb{L}^p$−norm of a real-valued random variable X, i.e. $\|X\|_p = \left[\mathbb{E}\left(X^p\right)\right]^{1/p}$.

Theorem 2. *Assume that Assumptions* **A1-A2** *hold true. Then the following conclusions hold true.*

1. *For any $x \in E$, the sequence $\left(f_{t-s}^t(x)\right)_{s\geq 0}$ converges almost surely and in $\mathbb{L}^p$ to a limit denoted by $X_t(x)$. Moreover, we have $\|d\left(f_{t-s}^t(x), X_t(x)\right)\|_p = O\left(\kappa^{s/m}\right)$.*
2. *For $x \neq y$, we have $\mathbb{P}\left(X_t(x) \neq X_t(y)\right) = 0$. We then set $X_t = X_t(x)$.*
3. *The process $(X_t)_{t\in\mathbb{Z}}$ is stationary and also ergodic if the process $(\zeta_t)_{t\in\mathbb{Z}}$ is itself ergodic.*
4. *If $(Y_t)_{t\in\mathbb{Z}}$ is a non-anticipative process (i.e. $Y_t \in \mathcal{F}_t$) such that $\sup_{t\in\mathbb{Z}} \mathbb{E}\left[d^p\left(Y_t, y_0\right)\right] < \infty$, then $Y_t = X_t$ a.s.*

Note that we do not consider the case $p \in (0,1)$, because in this case, one can simply replace the metric d by d^p which is again a metric.

Proof of Theorem 2 From Assumption **A1**, the triangular inequality and stationarity, it is clear that $\|d\left(f_t(x), x\right)\|_p$ is finite for any $x \in E$ and $t \in \mathbb{Z}$. Moreover, Assumption **A2** guarantees that for $(t, s, s') \in \mathbb{Z}^3$ such that $s' < s \leq t$ and $x, y \in E$, we have

$$\mathbb{E}\left[d^p\left(f_s^t(x), f_{s'}^t(y)\right)|\mathcal{F}_{t-1}\right] \leq L^p d^p\left(f_s^{t-1}(x), f_{s'}^{t-1}(x)\right).$$

We then have

$$\|d\left(f_s^t(x), f_{s'}^t(x)\right)\|_p \leq L\|d\left(f_s^{t-1}(x), f_{s'}^{t-1}(x)\right)\|_p. \tag{2}$$

With the same argument, when $s \leq t-m$, we also get

$$\|d\left(f_s^t(x), f_{s'}^t(x)\right)\|_p \leq \kappa\|d\left(f_s^{t-m}(x), f_{s'}^{t-m}(x)\right)\|_p.$$

We then show that for a given $x \in E$, $\|d\left(f_1^t(x), x\right)\|_p$ is finite. This follows from the following bounds.

$$\begin{aligned}\|d\left(f_1^t(x), f_1(x)\right)\|_p &\leq \sum_{i=1}^{t-1}\|d\left(f_{i+1}^t(x), f_i^t(x)\right)\|_p \\ &\leq \sum_{i=1}^{t-1} L^{t-i}\|d\left(x, f_1(x)\right)\|_p.\end{aligned}$$

1. For the first point, using Assumption **A2**, we have

$$\begin{aligned}\sum_{i\geq 0}\|d\left(f_{t-i}^t(x), f_{t-i-1}^t(x)\right)\|_p &\leq \sum_{i\geq 0}\kappa^{[(i+1)/m]}L^{i+1-[(i+1)/m]m}\|d\left(f_1(x), x\right)\|_p \\ &\leq \frac{L^m\kappa^{(1-m)/m}}{1-\kappa^{1/m}}.\end{aligned}$$

   This bound entails that the series $\sum_{i\geq 0} d\left(f_{t-i}^t(x), f_{t-i-1}^t(x)\right)$ is almost surely finite. By the Cauchy criterion, there exists a random variable $X_t(x)$ such that $\lim_{i\to\infty} d\left(f_{t-i}^t(x), X_t(x)\right) = 0$ a.s. The convergence also holds in $\mathbb{L}^p$, since from the Fatou lemma,

$$\begin{aligned}\|d\left(X_t(x), f_{t-i}^t(x)\right)\|_p &\leq \liminf_{j\to\infty}\|d\left(f_{t-j}^t(x), f_{t-i}^t(x)\right)\|_p \\ &\leq \sum_{s\geq i}\|d\left(f_{t-s}^t(x), f_{t-s-1}^t(x)\right)\|_p.\end{aligned}$$

   The second assertion then follows from the previous bounds.

2. If $x \neq y$, we have from the Fatou lemma and

$$\begin{aligned} \|d\left(X_t(x), X_t(y)\right)\|_p &\leq \liminf_{s\to\infty} \|d\left(f_{t-s}^t(x), f_{t-s}^t(y)\right)\|_p \\ &\leq \liminf_{s\to\infty} \kappa^{[(s+1)/m]} L^{s+1-[(s+1)/m]m} d(x,y) = 0. \end{aligned}$$

This shows the second point.

3. For the third point, we observe that for any $j \geq 1$ there exists a measurable function $H_j^{(x)}: E'^{j+1} \to E$ such that $f_{t-j}^t(x) = H_j^{(x)}\left(\zeta_t, \ldots, \zeta_{t-j}\right)$. Since $\lim_{j\to} H_j^{(x)}$ exists $\mathbb{P}_{(\zeta_{t-j})_{j\geq 0}}$ a.s. It is then possible to define a measurable function $H: (E')^{\mathbb{N}} \to \mathbb{R}$ such that $X_t = H\left((\zeta_{t-j})_{j\geq 0}\right)$ a.s. The process $(X_t)_{t\in\mathbb{Z}}$ has a Bernoulli shift structure with dependent entries and is then stationary and ergodic provided the process $(\zeta_t)_{t\in\mathbb{Z}}$ satisfies the same properties.

4. The last property follows from the following bounds which hold for any $j \geq 1$:

$$\begin{aligned} \|d(X_t, Y_t)\|_p &= \|d\left(f_{t-jm-1}^t(X_{t-mj}), f_{t-mj+1}(Y_{t-mj})\right)\|_p \\ &\leq \kappa^j \left[\|d(y_0, X_0)\|_p + \sup_{t\in\mathbb{Z}} \|d(y_0, Y_t)\|_p\right]. \end{aligned}$$

3 Functional dependence measure

The functional dependence measure has been introduced by Wu (2005) and is particularly interesting for autoregressive processes which are not necessarily strong mixing or for which getting strong mixing conditions requires additional regularity conditions on the noise distribution. The single requirement is to get a Bernoulli shift representation of the stochastic process of interest, i.e. $X_t = H\left(\xi_t, \xi_{t-1}, \ldots\right)$ where $(\xi_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables. Most of the limit theorems and deviation inequalities have been derived under such dependence measure. See for instance Wu (2007) and Wu and Wu (2016). Such asymptotic results have been applied to various statistical problems. See for instance Wu et al. (2010) for kernel estimation for time series, Xiao and Wu (2012) for covariance estimation or Liu and Wu (2010) for spectral density estimation. The notion of functional dependence is then an attractive alternative to the usual strong mixing when the process is defined by stochastic recursions. Our aim in this paper is to show that when the covariate process Z satisfies this kind of dependence, one can also control the functional dependence measure of the process X. We will then provide a new wide class of examples for which the aforementioned references provide an important number of statistical applications.

In this section, we assume that E is a subspace of $\mathbb{R}^k$ and the distance d is given by a norm $|\cdot|$ on $\mathbb{R}^k$. We assume that the random map f_t is given by a Bernoulli shift, i.e.

$$f_t(x) = H\left(x, \xi_t, \xi_{t-1}, \ldots\right), \quad t \in \mathbb{Z}, \quad x \in E,$$

where ξ is an a sequence of i.i.d. random variable taking values in a mesurable space $(G, \mathcal{G})$ and $H: E \times G^{\mathbb{N}} \to E$ is a measurable function. We then define a new sequence $\overline{\xi}$ such that $\overline{\xi}_0 = \xi_0'$ and $\overline{\xi}_t = \xi_t$ for $t \neq 0$. Moreover, for $t > 0$, let

$$\overline{f}_t(x) = H(x, \overline{\xi}_t, \overline{\xi}_{t-1}, \ldots).$$

we define for $t \geq 0$ and $p \geq 1$,

$$\theta_{p,t} = \mathbb{E}^{1/p}\left[\left|f_{-\infty}^{t} - \overline{f}_{\infty}^{t}\right|^{p}\right].$$

Moreover, for $h \in \mathbb{N}$, let $\Theta_{p,h} = \sum_{t \geq h} \theta_{p,t}$.

Our aim is to get an upper bound for the functional dependence coefficients $\Theta_{p,h}$. To this end, we add another assumption. Here we set for $t \in \mathbb{Z}$, $\mathcal{F}_t = \sigma\left(\xi_{t-j} : j \geq 0\right)$.

A4 There exists a measurable function $S : \mathbb{R}^d \to \mathbb{R}_+$ and $r, s \geq p$ such that $r^{-1} + s^{-1} = p^{-1}$ and $\mathbb{E}\left[S(f_{-\infty}^{0})^{s}\right] < \infty$ and for all $x \in E$ and $t \geq 1$,

$$\mathbb{E}^{1/p}\left[\left|\overline{f}_t(x) - f_t(x)\right|^{p} |\sigma(\xi_0') \vee \mathcal{F}_{t-1}\right] \leq S(x) H_{t-1},$$

where H_{t-1} is a random variable measurable with respect to $\sigma(\xi_0') \vee \mathcal{F}_{t-1}$ and such that $\mathbb{E}|H_{t-1}|^r < \infty$.

An immediate consequence of this assumption is that for any random variable V_{t-1}, measurable with respect to $\sigma(\xi_0') \vee \mathcal{F}_{t-1}$, we have

$$\|\overline{f}_t(V_{t-1}) - f_t(V_{t-1})\|_p \leq \|S(V_{t-1})\|_s \|H_{t-1}\|_r.$$

Proposition 1. *Assume that Assumptions* **A1-A4** *hold true. For any $h \geq 2$, there then exists $C_1 > 0$ not depending on h such that*

$$\Theta_{p,h} \leq C_1 \left[\kappa^{h/m} + \sum_{i=0}^{h-1} \kappa^{i/m} \eta_{r,h-i} + \sum_{i \geq h} \kappa^{i/m} \eta_{r,1}\right], \tag{3}$$

with

$$\eta_{r,j} = \sum_{t \geq j} \|H_{t-1}\|_r, \quad j \geq 1.$$

In particular, if $\eta_{r,1} < \infty$, there exists $C > 0$, not depending on h, such that

$$\Theta_{p,h} \leq C \left[\kappa^{h/m} + \sum_{i=0}^{h-1} \kappa^{i/m} \eta_{r,h-i}\right]. \tag{4}$$

Notes

1. From (4), we get a polynomial (resp. geometric) decrease of $(\Theta_{p,h})_{h \geq 1}$ as soon as $(\eta_{r,h})_{h \geq 1}$ has the same type of decrease.

2. Depending on the contraction property, it can happen that one can get a bound for $\theta_{p,t}$, while it is required a condition on $\theta_{q,t}$ or $\Theta_{q,h}$ for $q > p$ for applying some limit theorems or statistical results. This is still possible if $\mathbb{E}\left[|f_{-\infty}^{0}|^r\right] < \infty$ for some $r > q$. Indeed, from Hölder's inequality, we have

$$\theta_{q,t} \leq \theta_{p,t}^{\frac{p(r-q)}{q(r-p)}} \theta_{r,t}^{\frac{r(q-p)}{q(r-p)}}.$$

Moreover, $\theta_{r,t} \leq 2\|f_{-\infty}^{0}\|_r$.

Proof of Proposition 1 We use the decomposition

$$\begin{aligned}\overline{f}^t_{-\infty} - f^t_{-\infty} &= \sum_{i=0}^{t-1}\left[f^t_{t-i}\circ\overline{f}^{t-i-1}_{-\infty} - f^t_{t-i-1}\circ\overline{f}^{t-i-2}_{-\infty}\right]\\ &+ \overline{f}_t\circ\overline{f}^{t-1}_{-\infty} - f_t\circ\overline{f}^{t-1}_{-\infty}.\end{aligned}$$

From Assumption **A2** and Assumption **A4**, we have, for $i = 0, \ldots, t-2$,

$$\begin{aligned}\|f^t_{t-i}\circ\overline{f}^{t-i-1}_{-\infty} - f^t_{t-i-1}\circ\overline{f}^{t-i-2}_{-\infty}\|_p &\leq \kappa^{\frac{i+1}{m}-1}L^m\|\left(\overline{f}_{t-i-1} - f_{t-i-1}\right)\circ\overline{f}^{t-i-2}_{-\infty}\|_p\\ &\leq \kappa^{\frac{i+1}{m}-1}L^m\|S(\overline{f}^{t-i-2}_{-\infty})H_{t-i-2}\|_p\\ &\leq \kappa^{\frac{i+1}{m}-1}L^m\|S(f^0_{-\infty})\|_s\|H_{t-i-2}\|_r.\end{aligned}$$

If $i = t-1$, we have $\|f^t_{t-i}\circ\overline{f}^{t-i-1}_{-\infty} - f^t_{t-i-1}\circ\overline{f}^{t-i-2}_{-\infty}\|_p \leq 2\|f^0_{-\infty}\|_p\kappa^{t/m-1}$. Using the triangular inequality, we get for $t \geq 2$,

$$\theta_{p,t} \leq \kappa^{-1}L^m\|S(f^0_{-\infty})\|_s\sum_{i=0}^{t-2}\kappa^{(i+1)/m}\|H_{t-i-2}\|_r + \|S(f^0_{-\infty})\|_s\|H_{t-1}\|_r + 2\kappa^{t/m-1}\|f^0_{-\infty}\|_p.$$

The bound (3) is obtained by summation and entails the simpler bound (4).□

3.1 A result for general non linear autoregressions

Let $(\zeta_t)_{t\in\mathbb{Z}}$ be a stationary and ergodic processes taking values in a measurable space E' and adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$. Let also E be a subset of $\mathbb{R}^k$ and $|\cdot|$ a given norm on $\mathbb{R}^k$. Our aim is to study existence of solutions for the following recursive equations:

$$X_t = F\left(X_{t-1}, \ldots, X_{t-q}, \zeta_t\right), \quad t \in \mathbb{Z}, \tag{5}$$

where $F : E^q \times E' \to E$ is a measurable function. We first introduce additional notations. For any positive integer d, we denote by $\mathcal{M}_d$ the set of square matrices with real coefficients and d rows and if $A \in \mathcal{M}_d$, $\rho(A)$ the spectral radius of the matrix A. Moreover, for $x \in \mathbb{R}^d$ and $p \in \mathbb{R}_+$, the vector $(|x_1|^p, \ldots, |x_q|^p)'$ will be denoted by $|x|^p_{vec}$. Finally, we introduce a partial order relation $\preceq$ on $\mathbb{R}^d$ and such that $x \preceq x'$ means $x_i \leq x'_i$ for $i = 1, \ldots, d$.

We assume that the following conditions hold true for some real number $p \geq 1$.

B1 The process $(\zeta_t)_{t\in\mathbb{Z}}$ is stationary and ergodic.

B2 For any $y \in E^q$, $\mathbb{E}\left[|F(y, \zeta_1)|^p\right] < \infty$.

B3 There exist some matrices $A_1, \ldots, A_q \in \mathcal{M}_k$ with nonnegative elements, satisfying $\rho\left(A_1 + \cdot + A_q\right) < 1$ and such that for $y, y' \in E^q$,

$$\mathbb{E}\left[\left|F(y,\zeta_t) - F(y',\zeta_t)\right|^p_{vec}|\mathcal{F}_{t-1}\right] \preceq \sum_{i=1}^q A_i\left|y_i - y'_i\right|^p_{vec}.$$

Theorem 3. *Assume that Assumptions* **B1-B3** *hold true. There then exists a unique stationary and non-anticipative process* $(X_t)_{t\in\mathbb{Z}}$ *solution of* (5) *and such that* $\mathbb{E}\left[|X_t|^p\right] < \infty$. *Moreover, this process is ergodic.*

Proof of Theorem 3 Define the following random map

$$f_t(u_1,\dots,u_q)=\left(F\left(u_1,\dots,u_q,\zeta_t\right)',u_1',\dots,u_{q-1}'\right)'.$$

We set $x=(u_1,\dots,u_q)\in E^q$ and for $1\leq t\leq q$, $U_t(x)=u_{q-t+1}$. Next for $t\geq q+1$, we define $U_t(x)$ recursively by

$$U_t(x)=F\left(U_{t-1}(x),\dots,U_{t-q}(x),\varepsilon_t\right).$$

We then have for $t\geq q+1$,

$$(U_t(x),\dots,U_{t-q+1}(x))=f_{q+1}^t(x).$$

Using our assumptions, we have for $t\geq q+1$,

$$\mathbb{E}\left[|U_t(x)-U_t(x')|_{vec}^p|\mathcal{F}_{t-1}\right]\preceq\sum_{i=1}^q A_i\left|U_{t-i}(x)-U_{t-i}(x')\right|_{vec}^p.$$

We introduce the matrix

$$B=\begin{pmatrix}A_1 & \cdots & A_{q-1} & A_q\\ & I_{k(q-1)} & & 0_{k(q-1),1}\end{pmatrix}.$$

The condition $\rho(A_1+\cdots+A_q)<1$ entails that $\rho(B)<1$. Indeed, if $v=(v_1',\dots,v_q')'\in\mathbb{R}^{kq}\setminus\{0\}$ is such that $Bv=\lambda v$ for $|\lambda|\geq 1$, we get the equality $v_1=\left[\lambda^{-1}A_1+\cdots+\lambda^{-q}A_q\right]v_1$. Since the coefficients of the A_i's are nonnegative, we get

$$|v_1|_{vec}\preceq\sum_{i=1}^q|\lambda|^{-i}A_i|v_1|_{vec}\preceq\sum_{i=1}^q A_i|v_1|_{vec}.$$

We then get $|v_1|_{vec}\preceq\left(\sum_{j=1}^q A_j\right)^k|v_1|_{vec}$ any positive integer k. Letting $k\to\infty$, we obtain $v_1=0$. Since $v_i=\lambda v_{i+1}$ for $i=1,\dots,q-1$, we get $v=0$ which is a contradiction. Then $|\lambda|<1$ and $\rho(B)<1$. Next, we set

$$V_t(x)=\left(U_t(x)',\dots,U_{t-q+1}(x)'\right),\quad t\geq q+1.$$

Note that $V_t(x)=f_{q+1}^t(x)$. We then have

$$\mathbb{E}\left[|V_t(x)-V_t(x')|_{vec}^p|\mathcal{F}_{t-1}\right]\preceq B|V_{t-1}(x)-V_{t-1}(x')|_{vec}^p\preceq\cdots\preceq B^{t-q}|x-x'|_{vec}^p.$$

Setting for $v,v'\in\mathbb{R}^{kq}$, $d(v,v')=\left(\sum_{i=1}^{kq}|v_i-v_i'|^p\right)^{1/p}$, we get

$$\mathbb{E}\left[d^p\left(f_{q+1}^t(x),f_{q+1}^t(x')\right)|\mathcal{F}_q\right]\leq\left|\mathbb{1}'B^{t-q}\right|_\infty d^p(x,x'),$$

where $\mathbb{1}$ denotes the vector of $\mathbb{R}^{kq}$ having all its components equal to 1, and $|\cdot|_\infty$ the infinite norm in $\mathbb{R}^{kq}$. Since $\rho(B)<1$, if t is large enough, we have $\left|\mathbb{1}'B^{t-q}\right|_\infty<1$. We then conclude that **A2** is satisfied, with $m=\inf\{j\geq 1:\left|\mathbb{1}'B^j\right|_\infty<1\}$ and $\kappa^p=|\mathbb{1}'B^m|_\infty$. Moreover, **A1** is a direct consequence of **B2**. The result then follows from Theorem 2.$\square$

Next, we give additional assumptions for controlling the coefficients of functional dependence. More general conditions are possible but we give them for covariates that are exogenous in the sense discussed in the introduction. More precisely, we assume that $\zeta_t=(Z_{t-1},\varepsilon_t)$ where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a noise process taking values in the measurable space E' and $(Z_t)_{t\in\mathbb{Z}}$ is a covariate process taking values in $\mathbb{R}^d$.

B4 Let $(\eta_t)_{t\in\mathbb{Z}}$ be a sequence of random variables taking values in a measurable space $(G, \mathcal{G})$ and such that $Z_t = H(\eta_t, \eta_{t-1}, \ldots)$ for a measurable function H. Moreover, setting $\xi_t = (\varepsilon_t, \eta_t)$, we assume that $(\xi_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables. In this case, we assume that $\mathcal{F}_t = \sigma(\xi_s : s \leq t)$ for all $t \in \mathbb{Z}$.

B5 If $r > 0$ and $s \in \mathbb{R}_+ \cup \{\infty\}$ are such that $r^{-1} + s^{-1} = p^{-1}$, there exists a measurable function $S : E^q \to \mathbb{R}$ such that $\mathbb{E}[S(X_q, \ldots, X_1)^s] < \infty$ for all $z, z' \in \mathbb{R}^d$ and $x_1, \ldots, x_q \in E^q$,

$$\mathbb{E}^{1/p}\left[\left|F(x_1, \ldots, x_q, z, \varepsilon_0) - F(x_1, \ldots, x_q, z', \varepsilon_0)\right|^p\right] \leq S(x_1, \ldots, x_q)|z - z'|.$$

Note. Let us comment Assumption **B4**. Under this assumption, the η_t's are i.i.d. as well as the ε_t's and for any $t \in \mathbb{Z}$, ε_t is independent from $\mathcal{F}_{t-1} = \sigma(\xi_s : s \leq t-1\}$. Note that we allow simultanous dependence between ε_t and η_t. For instance, we can set $\eta_t = K(\varepsilon_t, U_t)$ where K is a mesurable function and U is a sequence of i.i.d. random variables, independent from the sequence ε. This assumption if then more flexible than the complete independence between the two error processes ε and η, which implies strict exogeneity.

The result is the following.

Proposition 2. *Assume that Assumptions* **B1** $-$ **B5** *hold true. There then exists* $C > 0$ *and* $\rho \in (0, 1)$*, such that for all* $h \geq 1$, $\Theta_{p,h}(X) \leq C\left[\rho^h + \sum_{i=1}^h \rho^i \Theta_{r,h-i}(Z)\right]$.

Proof of Proposition 2 Defining

$$f_t(x_1, \ldots, x_q) = (F(x_1, \ldots, x_q, Z_{t-1}, \varepsilon_t), x_1, \ldots, x_{q-1}),$$

we set for $t \geq 0$, $\overline{Z}_t = H(\eta_t, \ldots, \eta_1, \eta_0', \eta_{-1}, \ldots)$. Using Assumptions **B4-B5**, we have, for $t \geq 1$,

$$\mathbb{E}^{1/p}\left[\left|f_t(x_1, \ldots, x_q) - \overline{f}_t(x_1, \ldots, x_q)\right|^p |\mathcal{F}_{t-1} \vee \sigma(\xi_0')\right] \leq S(x_1, \ldots, x_q)|Z_{t-1} - \overline{Z}_{t-1}|.$$

We then check Assumption **A4**, setting $H_{t-1} = \left|Z_{t-1} - \overline{Z}_{t-1}\right|$. One can then apply Proposition 1 using for instance the ℓ_1-norm on $\mathbb{R}^q$. Since $\mathbb{E}^{1/r}\left[\left|Z_{t-1} - \overline{Z}_{t-1}\right|^r\right] = \theta_{r,t-1}(Z)$, the proof of the proposition is now complete.□

4 Examples

4.1 CHARME models

We consider the dynamic

$$Y_t = f(Y_{t-1}, \ldots, Y_{t-q}, Z_{t-1}) + \varepsilon_t g(Y_{t-1}, \ldots, Y_{t-q}, Z_{t-1}), \tag{6}$$

where q is a positive integer, $f, g : \mathbb{R}^q \times \mathbb{R}^d \to \mathbb{R}$ are measurable functions.

CH1 The process $((Z_t, \varepsilon_t))_{t\in\mathbb{Z}}$ is stationary and ergodic.

CH2 There exist measurable functions $a_i, b_i : \mathbb{R}^d \times \mathbb{R}_+$ such that

$$\left|f(y_1, \ldots, y_q, z) - f(y_1', \ldots, y_q', z)\right| \leq \sum_{i=1}^{q} a_i(z) \left|y_i - y_i'\right|,$$

$$\left|g(y_1, \ldots, y_q, z) - g(y_1', \ldots, y_q', z)\right| \leq \sum_{i=1}^{q} b_i(z) \left|y_i - y_i'\right|.$$

CH3 There exist a real number $p \geq 1$ such that $\|\varepsilon_1\|_p < \infty$ and $r, s \geq p$ such that $r^{-1} + s^{-1} = p^{-1}$, s can be infinite and two functions L_1, L_2 defined on $\mathbb{R}^q$ and such that for all $y_1, \ldots, y_q \in \mathbb{R}$ and $z, z' \in \mathbb{R}^d$,

$$\left|f(y_1, \ldots, y_q, z) - f(y_1, \ldots, y_q, z')\right| \leq L_1(y_1, \ldots, y_q)|z - z'|,$$

$$\left|g(y_1, \ldots, y_q, z) - g(y_1, \ldots, y_q, z')\right| \leq L_2(y_1, \ldots, y_q)\left|z - z'\right|.$$

Moreover, for $i = 1, 2$, $L_i(Y_q, \ldots, Y_1) \in \mathcal{L}_s$.

For $t \in \mathbb{Z}$ and $i = 1, \ldots, q$, we set $c_{i,t} = a_i\left(Z_{t-1}\right) + |\varepsilon_t| b_i\left(Z_{t-1}\right)|\varepsilon_t|$, We then define a sequence of random matrices $\mathbf{A} = (A_t)_{t\in\mathbb{Z}}$ by

$$A_t = \begin{pmatrix} c_{1,t} & c_{2,t} & \cdots & c_{q,t} \\ & & & 0 \\ & I_{q-1} & & \vdots \\ & & & 0 \end{pmatrix}$$

Finally, we denote by $\chi(\mathbf{A})$ the Lyapunov exponent of the sequence $\mathbf{A}$, i.e.

$$\chi(\mathbf{A}) = \lim_{n\to\infty} \frac{\mathbb{E}\left[\log \|A_n \cdots A_1\|\right]}{n},$$

where $\|\cdot\|$ is a norm on the space of square matrices of size $q \times q$.

Proposition 3. *Assume that Assumptions* **CH1-CH2** *hold true.*

1. *Suppose that $\chi(\mathbf{A}) < 0$. There then exists a unique non anticipative process $(Y_t)_{t\in\mathbb{Z}}$ solution of (6) which is stationary and ergodic.*

2. *Assume additionally that there exist $\overline{x} \in \mathbb{R}^q$, $p \geq 1$ such that $f(\overline{x}, Z_0) + \varepsilon_1 g(\overline{x}, Z_0) \in \mathbb{L}^p$ and some positive real numbers $\delta_1, \ldots, \delta_q > 0$ such that $\gamma := \sum_{i=1}^q \delta_i < 1$ and $\mathbb{E}^{1/p}\left[c_{i,1}^p|\mathcal{F}_0\right] \leq \delta_i$ a.s. There then exists a unique non anticipative process solution of (6) which is stationary, ergodic and such that $\mathbb{E}|Y_1|^p < \infty$.*

3. *Assume furthermore that Assumptions (***CH3***) and (***B4***) hold true. There then exists $C > 0$ and $\rho \in (0, 1)$, such that for all $h \geq 1$, $\Theta_{p,h}(X) \leq C\left[\rho^h + \sum_{i=1}^h \rho^i \Theta_{r,h-i}(Z)\right]$.*

Notes

1. For $q = 1$, the condition $\chi(\mathbf{A}) < 0$ reduces to $\mathbb{E}\left[\log\left(a_1(Z_0) + b_1(Z_0)|\varepsilon_1|\right)\right] < 0$. Note that the latter condition is much weaker than the condition $\mathbb{E}^{1/p}\left[\left(a_1(Z_0) + b_1(Z_0)|\varepsilon_1|\right)^p |\mathcal{F}_0\right] < \delta_1 < 1$ a.s., used in point 2. of Proposition 3. Indeed, the latter condition entails that

$$\mathbb{E}\left[a_1(Z_0) + b_1(Z_0)|\varepsilon_1|\right] \leq \|a_1(Z_0) + b_1(Z_0)|\varepsilon_1|\|_p < 1$$

   and from Jensen's inequality, $\log \mathbb{E}\left[a_1(Z_0) + b_1(Z_0)|\varepsilon_1|\right] \leq \chi(\mathbf{A})$.

   For $q \geq 2$, it is more difficult to obtain explicit conditions which guaranty that $\chi(\mathbf{A}) < 0$. Note that if ε_t and $\mathcal{F}_{t-1}$ are independent for all $t \in \mathbb{Z}$ (weak exogeneity condition), the condition $\sum_{i=1}^q \left[\sup_z a_i(z) + \sup_z b_i(z)\|\varepsilon_1\|_p\right] < 1$. for the additional condition given in point 2.

2. Using the results of Liu et al. (2010), a nonparametric kernel estimation of the functions f and g is possible. Proposition 3 gives precise assumptions under which it is possible to apply their results when the regressors in f and g include lag values of the response and we then obtain additional examples of time series models for which this kind of nonparametric estimation is possible.

Proof of Proposition 3

1. Let d be the distance induced by the ℓ_1-norm on $\mathbb{R}^q$, i.e. $d(x, y) = \sum_{i=1}^q |x_i - y_i|$. We use the notation $|x - y|$ instead of $d(x, y)$. For a square matrix A of size $q \times q$, we denote by $\|A\|$ the corresponding operator norm of A. We define the sequence of random maps as follows:

$$f_t(x) = \left(f(x_1, \ldots, x_q, Z_{t-1}) + \varepsilon_t g(x_1, \ldots, x_q, Z_{t-1}), x_1, \ldots, x_{q-1}\right)'.$$

   We then have

$$|f_t(x) - f_t(y)|_{vec} \preceq \left(\sum_{i=1}^q c_{i,t}|x_i - y_i|, |x_1 - y_1|, \ldots, |x_{q-1} - y_{q-1}|\right)' = A_t \cdot |x - y|_{vec}.$$

   Iterating the previous bound, we get for any positive integer t,

$$\left|f_1^t(x) - f_1^t(y)\right|_{vec} \preceq A_t \cdots A_1 \cdot |x - y|_{vec}.$$

   We then deduce that $c(f_1^t) \leq \|A_t \cdots A_1\|$. The result is then a consequence of Theorem 1, using the condition $\chi(\mathbf{A}) < 0$.

2. We check the assumptions of Theorem 3. First note that from Assumption **CH1**, the process $\zeta_t = (Z_{t-1}, \varepsilon_t)$ is ergodic. This entails **B1**. Next we set

$$F(x_1, \ldots, x_q, \zeta_t) = f(x_1, \ldots, x_q, Z_{t-1}) + \varepsilon_t g(x_1, \ldots, x_q, Z_{t-1}).$$

   Our assumptions guaranty that $F(\overline{x}, \zeta_1) \in \mathbb{L}^p$ and, using **CH2**, we deduce that $F(x, \zeta_1) \in \mathbb{L}^p$ for any $x \in \mathbb{R}^q$. This shows **B2**. Finally, we check **B3**. Using Minkowski's inequality for conditional expectations (see for instance Doob (2012), Chapter XI, Section 3), we have

$$\begin{aligned}
\mathbb{E}^{1/p}\left[|F(x_1,\ldots,x_q,\zeta_t)-F(y_1,\ldots,y_q,\zeta_t)|^p\,|\mathcal{F}_{t-1}\right] &\leq \mathbb{E}^{1/p}\left[\left(\sum_{i=1}^q c_{i,t}|x_i-y_i|\right)^p|\mathcal{F}_{t-1}\right]\\
&\leq \sum_{i=1}^q \mathbb{E}^{1/p}\left[c_{i,t}^p|\mathcal{F}_{t-1}\right]\cdot|x_i-y_i|\\
&\leq \sum_{i=1}^q \delta_i|x_i-y_i|.
\end{aligned}$$

Next using convexity, we get

$$\mathbb{E}\left[|F(x_1,\ldots,x_q,\zeta_t)-F(y_1,\ldots,y_q,\zeta_t)|^p\,|\mathcal{F}_{t-1}\right]\leq\left(\sum_{i=1}^q\delta_i\right)^{p-1}\sum_{i=1}^q\delta_i|x_i-y_i|.$$

B3 is then a consequence of **CH4**.

3. We apply Proposition 2. From the previous points, it is only required to check **B5** which is a consequence of Assumption **CH4**.□

4.2 GARCH processes

GARCH processes with exogenous regressors have been considered recently by Pedersen and Rahbek (2018) or Francq and Thieu (2019). We consider here the asymmetric power GARCH studied by Francq and Thieu (2019). The model is defined as follows.

$$Y_t=\varepsilon_t h_t^{1/\delta},\quad h_t=\pi' Z_{t-1}+\sum_{i=1}^q\left\{\beta_i h_{t-i}+\alpha_{i+}(Y_{t-i}^+)^\delta+\alpha_{i-}(Y_{t-i}^-)^\delta\right\},\tag{7}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ and $(Z_t)_{t\in\mathbb{Z}}$ are two sequences of random variables taking values in $\mathbb{R}$ and $\mathbb{R}_+^d$ respectively, $\delta>0$, $\pi\in\mathbb{R}_+^d$ and the β_i's, α_{i+}'s and α_{i-}'s are nonnegative real numbers. Optimal stationarity properties of time series models defined by (7) have been obtained by Francq and Thieu (2019), using a version of Theorem 1 for affine random maps. In contrast, we use our results to get existence of a moment of order δ for the unique stationary solution. The following assumptions will be needed.

G1 The process $((Z_t,\varepsilon_t))_{t\in\mathbb{Z}}$ is stationary and ergodic and $\mathbb{E}|Z_0|<\infty$.

G2 There exist s_-,s_+ such that $\mathbb{E}\left[(\varepsilon_t^+)^\delta|\mathcal{F}_{t-1}\right]\leq s_+$ and $\mathbb{E}\left[(\varepsilon_t^-)^\delta|\mathcal{F}_{t-1}\right]\leq s_-$ a.s. and $\gamma:=\sum_{i=1}^p(\beta_i+s_+\alpha_{i+}+s_-\alpha_{i-})<1$.

Proposition 4. *Assume that Assumptions* **G1-G2** *hold true.*

1. *There then exists a unique stationary and non anticipative solution $(Y_t)_{t\in\mathbb{Z}}$ solution of (7). This solution is ergodic and satisfies $\mathbb{E}|Y_0|^\delta<\infty$.*
2. *Additionally, assume that Assumption* **B4** *holds true. Let $H_t=\left((Y_t^+)^\delta,(Y_t^+)^\delta,h_t\right)$. There then exists $C>0$ and $\rho\in(0,1)$, such that for all $h\geq 1$, $\Theta_{1,h}(H)\leq C\left[\rho^h+\sum_{i=1}^h\rho^i\Theta_{1,h-i}(Z)\right]$.*

Note. When $\delta = 2$, $\alpha_{j+} = \alpha j- = \alpha_j$ and ε_t is independent from $\sigma\left((\varepsilon_s, Z_s) : s \leq t-1\right)$, the contraction condition in **G2** reduces to $\mathbb{E}\left(\varepsilon_0^2\right)\sum_{j=1}^q(\alpha_j+\beta_j) < 1$, which is the classical condition ensuring existence of a standard GARCH process with finite second moment.

Proof of Proposition 4 For the first part, we apply Proposition 3. To this end, we set $E = \mathbb{R}_+^3$, $F = (F_1, F_2, F_3)$, $F_2(y_1, \ldots, y_q, \zeta_t) = (\varepsilon_t^+)^\delta y_{1,1}$, $F_3(y_1, \ldots, y_q, \zeta_t) = (\varepsilon_t^-)^\delta y_{1,1}$ and

$$F_1\left(y_1, \ldots, y_q, \zeta_t\right) = \pi' Z_{t-1} + \sum_{j=1}^q \beta_j y_{1,j} + \left(\alpha_{1+}(\varepsilon_t^+)^\delta + \alpha_{1-}(\varepsilon_t^-)^\delta\right) y_{1,1} + \sum_{j=2}^q \left(\alpha_{j+} y_{2,j} + \alpha_{j-} y_{3,j}\right).$$

We then deduce that Assumption **B3** holds true with

$$A_1 = \begin{pmatrix} \beta_1 + \alpha_{1+}s_+ + \alpha_{1-}s_- & 0 & 0 \\ s_+ & 0 & 0 \\ s_- & 0 & 0 \end{pmatrix}, \quad A_j = \begin{pmatrix} \beta_j & \alpha_{j+} & \alpha_{j-} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad j \geq 2.$$

It is straightforward to show that the matrix $\Gamma := \sum_{j=1}^q A_j$ has eigenvalues 0 and $\frac{a \pm \sqrt{a^2+4(bd+ce)}}{2}$ with $a = \sum_{j=1}^q \beta_j + \alpha_{1+}s_+ + \alpha_{1-}s_-$, $b = \sum_{j=2}^q \alpha_{j+}$, $c = \sum_{j=2}^q \alpha_{j-}$, $d = s_+$ and $e = s_-$. Condition $\rho(\Gamma) < 1$ is equivalent to $\gamma < 1$ It is then clear that **B1-B3** follow from **G1-G2**.

For the second part, it is easily seen that **B5** is satisfied for a constant function S, $p = 1$ and $s = \infty$.□

4.3 Poisson autoregressions

We consider the PARX model introduced in Agosto et al. (2016). The idea is to model the conditional distribution of Y_t given $\mathcal{F}_{t-1}$ by a Poisson distribution with a random intensity λ_t depending on past values and a covariate process. More precisely, we assume that

$$Y_t = N_{\lambda_t}^{(t)}, \quad \lambda_t = \beta_0 + \sum_{j=1}^q \beta_j \lambda_{t-j} + \sum_{j=1}^q \alpha_j Y_{t-j} + \pi' Z_{t-1}, \tag{8}$$

where $\left(N^{(t)}\right)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. Poisson processes with intensity 1, $\beta_0, \ldots, \beta_q$, $\alpha_1, \ldots, \alpha_q$ and nonnegative real numbers and π is a vector of $\mathbb{R}^d$ with nonnegative coordinates.

PA1 We have $\gamma := \sum_{j=1}^q \alpha_j + \sum_{j=1}^q \beta_j < 1$.

PA2 The process $\left((Z_t, N^{(t)})\right)_{t\in\mathbb{Z}}$ is stationary, ergodic and adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ such that for all $t \in \mathbb{Z}$, $N^{(t)}$ is independent from $\mathcal{F}_{t-1}$. Moreover, $\mathbb{E}|Z_1| < \infty$.

PA3 There exists a sequence $(\eta_t)_{t\in\mathbb{Z}}$ of i.i.d. random variables, taking values in a measurable space G and a measurable function $H : G^{\mathbb{Z}} \to \mathbb{R}^d$ such that $Z_t = H\left(\eta_t, \eta_{t-1}, \ldots\right)$ a.s.

Proposition 5. *1. Assume that Assumptions* **PA1-PA2** *hold true. There then exists a unique non anticipative, stationary and ergodic process $(Y_t)_{t\in\mathbb{Z}}$ solution of (8).*

2. Additionally, if Assumptions **PA3** *and* **B4** *are also satisfied, with $\varepsilon_t = N^{(t)}$, there then exists $C > 0$ and $\rho \in (0,1)$, such that for all $h \geq 1$, $\Theta_{1,h}\left((Y_t, \lambda_t)_t\right) \leq C\left[\rho^h + \sum_{i=1}^h \rho^i \Theta_{1,h-i}(Z)\right]$.*

Note. Our result extends substantially that of Agosto et al. (2016). First, we prove ergodicity properties in PARX models without assuming that the covariate process $(Z_t)_{t\in\mathbb{Z}}$ is a Markov chain defined by a random map contracting in average. Secondly, for the stochastic dependence properties, we control the coefficient of functional dependence measure only assuming a general Bernoulli shift representation for $(Z_t)_{t\in\mathbb{Z}}$. For instance, $(Z_t)_{t\in\mathbb{Z}}$ can be defined by an infinite moving average process and is not necessarily Markovian.

Proof of Proposition 5

1. To show the first point, we check the assumptions of Theorem 3. We set $\zeta_t = \left(N^{(t)}, Z_{t-1}\right)$, $E = \mathbb{N}\times\mathbb{R}_+$ and the state space E is endowed with the ℓ_1-norm. We first note that $(Y_t)_{t\in\mathbb{Z}}$ is a stationary solution of (8) if and only if $X_t = (Y_t, \lambda_t)'$ is solution of $X_t = F\left(X_{t-1}, \dots, X_{t-q}, \zeta_t\right)$ with
$$F\left(x_1, \dots, x_q, \zeta_t\right) = \left(N^{(t)}_{f(x_1,\dots,x_q,Z_{t-1})}, f(x_1, \dots, x_q, Z_{t-1})\right)'$$
and $f(x_1, \dots, x_q, Z_{t-1}) = \beta_0 + \sum_{j=1}^q \beta_j s_j + \sum_{j=1}^q \alpha_j y_j + \pi' Z_{t-1}$, $x_i = (y_i, s_i)$, $1 \le i \le q$. For $x \in (\mathbb{N}\times\mathbb{R}_+)^q$,
$$\mathbb{E}\left[|F\left(x, \zeta_1\right)|\right] = 2\left(\beta_0 + \sum_{j=1}^q \beta_j s_j + \sum_{j=1}^q \alpha_j y_j + \pi'\mathbb{E}(Z_1)\right) < \infty$$
since $\mathbb{E}(|Z_1|) < \infty$.
We then have, for $(x, x') \in \left((\mathbb{N}\times\mathbb{R}_+)^q\right)^2$ with $x = (x_1, \dots, x_q), x' = (x'_1, \dots, x'_q)$ $\forall j = 1, \dots, q, x_j = (y_j, s_j), x_j = (y'_j, s'_j)$,
$$\mathbb{E}\left[\left|F(x, \zeta_t) - F(x', \zeta_t)\right|_{vec} |\mathcal{F}_{t-1}\right] \preceq \sum_{j=1}^q \begin{pmatrix}\alpha_j & \beta_j\\ \alpha_j & \beta_j\end{pmatrix} \left|x_j - x'_j\right|_{vec}.$$
In the previous bounds, we have used the identity $\mathbb{E}\left[|N^{(t)}_{h_{t-1}} - N^{(t)}_{g_{t-1}}||\mathcal{F}_{t-1}\right] = |h_{t-1} - g_{t-1}|$ which is valid for two nonnegative random variables h_{t-1}, g_{t-1} measurable with respect to $\mathcal{F}_{t-1}$. The previous equality follows from the properties of the Poisson process. Letting
$$\Gamma = \sum_{j=1}^q \begin{pmatrix}\alpha_j & \beta_j\\ \alpha_j & \beta_j\end{pmatrix},$$
The matrix Γ has two eigenvalues: 0 and γ. Assumption **PA1** then guarantees that $\rho(\Gamma) < 1$. Assumptions **B1-B3** of Theorem 3 are satisfied. Hence, according to Theorem 3, there exists a unique stationary and non-anticipative process $(X_t)_{t\in\mathbb{Z}}$ solution of (8) and such that $\mathbb{E}\left[|X_t|\right] < \infty$. This process is ergodic. This completes the proof of the first point.

2. For the second point, we use Proposition 2. To this end, it is only necessary to check **B5** for $p = r = 1$ and $s = \infty$. This is straightforward since we have the equality
$$\mathbb{E}\left[|F(x_1, \dots, x_q, z, N^{(t)}) - F(x_1, \dots, x_q, \overline{z}, N^{(t)})|\right] = 2|\pi'(z - \overline{z})|.$$
The proof of the second point is now complete.□

4.4 Dynamic binary choice model

We consider the dynamic

$$Y_t = \mathbb{1}_{g(Y_{t-1},\ldots,Y_{t-q},\zeta_t)>0}, \tag{9}$$

Where $(\zeta_t)_{t\in\mathbb{Z}}$ is a stationary process taking values in a measurable space E' and $g : \{0,1\}^q \times E' \to \mathbb{R}$ is a measurable function. This kind of binary model is popular in econometrics for studying dynamic of recessions. See de Jong and Woutersen (2011) who studied the case g linear and Kauppi and Saikkonen (2008) for a study of US recessions.

Proposition 6. *1. Assume that $(\zeta_t)_{t\in\mathbb{Z}}$ is a stationary and ergodic process such that*

$$\mathbb{P}\left(\min_{y\in\{0,1\}^q, 1\le t\le q} g(y,\zeta_t) > 0\right) + \mathbb{P}\left(\max_{y\in\{0,1\}^q, 1\le t\le q} g(y,\zeta_t) \le 0\right) > 0. \tag{10}$$

There then exists a unique stationary and ergodic solution $(Y_t)_{t\in\mathbb{Z}}$ for the recursions (9).

2. *Assume that for some real numbers $a_1,\ldots,a_q$ and $\pi\in\mathbb{R}^d$, $g(y,\zeta_t) = \sum_{i=1}^q a_i y_i + \pi' Z_{t-1} + \varepsilon_t$, with $\zeta_t = (Z_{t-1},\varepsilon_t)$ satisfying* **B4** *and the c.d.f. F_ε of ε_t being Lipschitz and taking values in $(0,1)$. Moreover, setting $\upsilon_t = \pi' Z_{t-1} + \varepsilon_t$, we assume that there exists $\delta > 0$ and a positive integer K such that*

$$\mathbb{P}\left(\phi_- + \min_{1\le t\le q} \upsilon_t > 0 | \mathcal{F}_{-K}\right) + \mathbb{P}\left(\phi_+ + \max_{1\le t\le q} \upsilon_t \le 0 | \mathcal{F}_{-K}\ldots\right) \ge \kappa \ a.s., \tag{11}$$

where

$$\phi_+ = \max\left\{\sum_{i=1}^q a_i y_i : (y_1,\ldots,y_n)\in\{0,1\}^n\right\}, \quad \phi_- = \min\left\{\sum_{i=1}^q a_i y_i : (y_1,\ldots,y_n)\in\{0,1\}^n\right\}.$$

There then exists $C > 0$ and $\rho\in(0,1)$, such that for all $h\ge 1$,

$$\Theta_{1,h}(Y) \le C\left[\rho^h + \sum_{i=1}^h \rho^i \Theta_{1,h-i}(Z)\right].$$

Notes

1. de Jong and Woutersen (2011) derived existence of a unique stationary and ergodic solution for (9) under the condition (11). As shown in de Jong and Woutersen (2011), Condition (11) holds in particular when the process $(\upsilon_t)_{t\in\mathbb{Z}}$ is $m-$dependent or for some infinite moving averages. Condition (10) is weaker. As the proof of Proposition 6 will show, the condition (11) implies Assumption **A2**. Condition (10) is only based on Theorem 1 and the Lyapunov exponent of some sequence of random maps. However, (10) does not entail mixing properties while condition (11) yes. See de Jong and Woutersen (2011), Theorem 2. Our results (see point 2. of Proposition 6) give a complement when the covariate process is not necessarily strongly mixing and has a Bernoulli shift representation.

2. When $\zeta_t = (Z_{t-1}, \varepsilon_t) \in \mathbb{R}^{d+1}$ in (9), one can allow interactions between lag values of the response and the covariates. For example,

$$g(y, \zeta_t) = \sum_{i=1}^{d} c_i y_i + \sum_{i=1}^{q} \sum_{j=1}^{d} \left[a_{i,j} y_i + b_{i,j}(1 - y_i)\right] Z_{j,t-i} + \varepsilon_t.$$

When ε_t is independent of $\mathcal{F}_{t-1} = \sigma\left((\varepsilon_{t-j}, Z_{t-j}) : j \geq 1\right)$, condition (10) is satisfied as soon as the distribution of ε_t has a support equal to the whole real line. We will not give a control of the functional dependence measure for this model because we were not able to check **A2** when the covariate process $(X_t)_{t\in\mathbb{Z}}$ is not bounded. However when the cdf of ε_t is known (e.g. for the logistic or the probit model), it is widely known that ergodicity of the process is sufficient for showing consistency and asymptotic normality of conditional pseudo likelihood estimators of the parameters.

Proof of Proposition 6

1. We apply Theorem 1. To this end, we define the random map from $E = \{0, 1\}^q$ to E by

$$f_t(x) = \left(\mathbb{1}_{g(x,\zeta_t)>0}, x_1, \ldots, x_{q-1}\right)'.$$

We set

$$\delta_t := \max_{y,y'\in\{0,1\}^q} \left|\mathbb{1}_{g(y,\zeta_t)>0} - \mathbb{1}_{g(y',\zeta_t)>0}\right| \leq \mathbb{1}_{\max_{y\in\{0,1\}^q} g(y,\zeta_t)>0} - \mathbb{1}_{\min_{y\in\{0,1\}^q} g(y,\zeta_t)>0}.$$

Setting $(y_t(x), \ldots, y_{t-q+1}(x))' = f_1^t(x)$ for $t \geq q$, we have

$$y_t(x) = 1 \text{ if and only if } g\left(y_{t-1}(x), \ldots, y_{t-q}(x), \zeta_t\right) > 0.$$

We have, setting $c_t = |y_t(x) - y_t(x')|$,

$$c_t \leq \delta_t \max_{1\leq j\leq q} c_{t-j}.$$

Using the fact that $\delta_t \leq 1$, a straightforward induction on $i = 0, \ldots, q-1$ shows that

$$c_{t+i} \leq \delta_{t+i} \max_{1\leq j\leq q} c_{t-j}.$$

Setting $d(y, y') = \max_{1\leq i\leq q} |y_i - y'_i|$,

$$d\left(f_1^{t+q-1}(x), f_1^{t+q-1}(x')\right) \leq \max_{0\leq i\leq q-1} \delta_{t+i} d\left(f_1^{t-1}(x), f_1^{t-1}(x')\right).$$

This shows in particular that

$$c\left(f_1^q\right) \leq \max_{1\leq i\leq q} \delta_i \leq \mathbb{1}_{\max_{y,i} g(y,\zeta_i)>0} - \mathbb{1}_{\min_{y,i} g(y,\zeta_i)>0}. \tag{12}$$

From our assumptions, the last upper bound can vanish with positive probability, and then $\mathbb{E}\left[\log c(f_1^q)\right] = -\infty = \chi$. Theorem 1 leads to the result.

2. The result will follow from Proposition 1. To this end, we check Assumptions **A1-A3**. **A1** is automatic we use the metric d on $\{0,1\}^q$ which is bounded. We set $h_t = \max_{t-q+1\le i\le t}\delta_i$. Note that in the linear case, we have

$$h_t \le \mathbb{1}_{\phi_+ + \max_{t-q+1\le i\le t} \upsilon_i>0} - \mathbb{1}_{\phi_- + \min_{t-q+1\le i\le t}\upsilon_i>0}.$$

To check **A2**, we use the bound (12) and the inequality $h_s \le 1$ for all $s\in\mathbb{Z}$ to get

$$c\left(f_{t-Jq+1}^t\right) \le \prod_{j=0}^{J-1} h_{t-jq} \le h_t.$$

Moreover using (11), we have

$$\begin{aligned}\mathbb{E}\left(h_t|\mathcal{F}_{t-Jq}\right) &\le 1-\mathbb{P}\left(\phi_+ + \max_{t-q+1\le i\le t}\upsilon_i \le 0|\mathcal{F}_{t-Jq}\right) - \mathbb{P}\left(\phi_- + \min_{t-q+1\le i\le t}\upsilon_i \ge 0|\mathcal{F}_{t-Jq}\right)\\ &\le 1-\kappa,\end{aligned}$$

provided that $Jq \ge q+K$. This guarantees **A2**, with $\kappa = 1-\delta$ and $m = Jq$.

Finally, let us check **A3**. Setting $\overline{Z}_t = H\left(\eta_t,\ldots,\eta_1,\eta_0',\eta_{-1},\ldots\right)$. We have

$$\begin{aligned}\mathbb{E}\left[d\left(f_t(y),\overline{f}_t(y)\right)|\mathcal{F}_{t-1}\vee\sigma(\xi_0')\right] &\le \left|F_\varepsilon\left(-\sum_{i=1}^q a_i y_i - \pi' Z_{t-1}\right) - F_\varepsilon\left(-\sum_{i=1}^q a_i y_i - \pi'\overline{Z}_{t-1}\right)\right|\\ &\le L_\varepsilon \cdot \max_{j=1}^d |\pi_j| \cdot \sum_{j=1}^d \left|Z_{j,t-1} - \overline{Z}_{j,t-1}\right|,\end{aligned}$$

where L_ε denotes the Lipschitz constant of F_ε.□

4.5 Categorical time series with covariates

We consider a finite set $E=\{1,2,\ldots,N\}$, an integer $q\ge 1$, a process $(Z_t)_{t\in\mathbb{Z}}$ taking values in $\mathcal{Z}\subset\mathbb{R}^d$ and a family $\{K_z(\cdot|\cdot): z\in\mathcal{Z}\}$ of probability kernels from E^q to E. Our aim is to construct a process $(Y_t)_{t\in\mathbb{Z}}$, taking values in E and such that

$$\mathbb{P}\left(Y_t = i|Y_{t-1}^-, Z_{t-1}^-\right) = K_{Z_{t-1}}\left(i|Y_{t-1},\ldots,Y_{t-q}\right).$$

A particular example is given by the multinomial autoregression, i.e.

$$K_z(i,y_1,\ldots,y_q) = \frac{\exp\left(\sum_{j=1}^q a_{i,j}y_j + \gamma_i' z\right)}{\sum_{k=1}^N \exp\left(\sum_{j=1}^q a_{k,j}y_j + \gamma_k' z\right)}$$

and is a classical model for categorical time series. See Fokianos and Kedem (2003). In econometrics, Russell and Engle (2005) studied the dynamic of price changes using such model but with a more general observation-driven form as in GARCH models that will not fall in our framework.

For applying our results, we now define some random maps. For $t \in \mathbb{Z}$, let ε_t be a random variable uniformly distributed over $[0,1]$. For $u \in [0,1]$, $z \in \mathcal{Z}$, $y \in E^q$ and $u \in [0,1]$, we set

$$K_z^-(u|y) = \inf\left\{ i = 1, \ldots, N : \sum_{j=1}^{i} K_z(j|y) \geq u \right\}$$

and

$$f_t(y_1, \ldots, y_q) = \left(K_{Z_{t-1}}^- \left(\varepsilon_t | y_1, \ldots, y_q\right), y_1, \ldots, y_{q-1} \right)'.$$

We introduce the following assumptions.

C1 The probability kernels K_z are lower bounded by a positive constant not depending on z, i.e. $\eta := \inf_{z \in \mathcal{Z}, (i,y) \in E^{q+1}} K_z(i|y) > 0$.

C2 For $t \in \mathbb{Z}$, ε_t is independent from $\mathcal{F}_{t-1} = \sigma\left((Z_s, \varepsilon_s) : s \leq t-1\right)$. Moreover, the process $(Z_t)_{t \in \mathbb{Z}}$ is stationary and ergodic.

C3 There exists a constant $C > 0$ such that for all $y_1, \ldots, y_q \in E$,

$$\sum_{i=1}^{N} |K_z(i|y_1, \ldots, y_q) - K_{\overline{z}}(i|y_1, \ldots, y_q)| \leq C|z - \overline{z}|.$$

Theorem 4. *Assume that Assumptions* **C1-C2** *hold true.*

1. *The conclusions of Theorem 2 hold true for the discrete metric $d(y, y') = \mathbb{1}_{y \neq y'}$. As a consequence, there exists a unique stationary and non anticipative process $(Y_t)_{t \in \mathbb{Z}}$ such that*

$$Y_t = K_{Z_{t-1}}^- \left(\varepsilon_t | Y_{t-1}, \ldots, Y_{t-q}\right), \quad t \in \mathbb{Z}.$$

   The process $((Y_t, Z_t))_{t \in \mathbb{Z}}$ is also ergodic.

2. *Additionally, assume that Assumption* **B4** *and* **C3** *hold true. There then exists $C > 0$ and $\rho \in (0,1)$, such that for all $h \geq 1$, $\Theta_{1,h}(Y) \leq C\left[\rho^h + \sum_{i=1}^{h} \rho^i \Theta_{1,h-i}(Z)\right]$.*

Note. For the multinomial autoregressive model, **C1** is only satisfied when $(Z_t)_{t \in \mathbb{Z}}$ is a bounded process. Recently Fokianos and Truquet (2019) studied the same model under the strict exogeneity assumption with an unbounded covariate process. However, we do not assume here that the two processes $(Z_t)_{t \in \mathbb{Z}}$ and $(\varepsilon_t)_{t \in \mathbb{Z}}$ are independent which is less restrictive than strict exogeneity.

Proof Theorem 4

1. We apply Theorem 2. Since d is a bounded metric, **A1** is automatically satisfied. Next, setting for $x \in E^q$, $X_t(x) = f_t \circ \cdots \circ f_{t-q+1}(x)$, we have

$$\begin{aligned} \mathbb{P}\left(X_t(x) = X_t(x') | \mathcal{F}_{t-q}\right) &\geq \mathbb{P}\left(X_t(x) = X_t(x') = (1, \ldots, 1) | \mathcal{F}_{t-q}\right) \\ &\geq \mathbb{P}\left(\varepsilon_t, \ldots, \varepsilon_{t-q+1} \in [0, \eta]\right) \\ &\geq \eta^q. \end{aligned}$$

   This yields to the bound $\mathbb{E}\left[d\left(X_t(x), X_t(x') | \mathcal{F}_{t-q}\right)\right] \leq 1 - \eta^q$, which shows the second part of **A2**. The first part is automatic. The result is then a consequence of Theorem 2.

2. For the second part, we check **A3**. Note that for $i, j \in E$, we have $(N-1)^{-1}|i-j| \leq \mathbb{1}_{i \neq j} \leq |i-j|$. Using the ℓ_1-metric on E^q which is equivalent to the discrete metric, we have

$$\begin{aligned}
\mathbb{E}\left[|f_t(x) - \overline{f}_t(x)| \middle| \mathcal{F}_{t-1} \vee \sigma(\xi_0')\right] &\leq \mathbb{E}\left[|K^-_{Z_{t-1}}(\varepsilon_t|y_1,\ldots,y_q) - K^-_{\overline{Z}_{t-1}}(\varepsilon_t|y_1,\ldots,y_q)| \middle| \mathcal{F}_{t-1} \vee \sigma(\xi_0')\right] \\
&\leq \int_0^1 \left|K^-_{Z_{t-1}}(u|y_1,\ldots,y_q) - K^-_{\overline{Z}_{t-1}}(u|y_1,\ldots,y_q)\right| du \\
&\leq \sum_{j=1}^N \left|\sum_{i=1}^j K_{Z_{t-1}}(i|y_1,\ldots,y_q) - \sum_{i=1}^j K_{\overline{Z}_{t-1}}(i|y_1,\ldots,y_q)\right| \\
&\leq NC\left|Z_{t-1} - \overline{Z}_{t-1}\right|.
\end{aligned}$$

Assumption **A3** then follows with $s = \infty$ and $r = p = 1$.□

References

Arianna Agosto, Giuseppe Cavaliere, Dennis Kristensen, and Anders Rahbek. Modeling corporate defaults: Poisson autoregressions with exogenous covariates (parx). *Journal of Empirical Finance*, 38:640–663, 2016.

Aleksandr Alekseevich Borovkov. Ergodicity and stability of stochastic processes. 1998.

Philippe Bougerol and Nico Picard. Strict stationarity of generalized autoregressive processes. *The Annals of Probability*, pages 1714–1730, 1992.

Andreas Brandt. The stochastic equation y n+ 1= a n y n+ b n with stationary coefficients. *Advances in Applied Probability*, 18(1):211–220, 1986.

Gary Chamberlain. The general equivalence of granger and sims causality. *Econometrica: Journal of the Econometric Society*, pages 569–581, 1982.

Richard A Davis and Heng Liu. Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, pages 1673–1707, 2016.

Robert M de Jong and Tiemen Woutersen. Dynamic time series binary choice. *Econometric Theory*, 27(4):673–702, 2011.

Persi Diaconis and David Freedman. Iterated random functions. *SIAM review*, 41(1):45–76, 1999.

Joseph L Doob. *Measure theory*, volume 143. Springer Science & Business Media, 2012.

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

John H Elton. A multiplicative ergodic theorem for lipschitz maps. *Stochastic Processes and their Applications*, 34(1):39–47, 1990.

Robert F Engle, David F Hendry, and Jean-Francois Richard. Exogeneity. *Econometrica: Journal of the Econometric Society*, pages 277–304, 1983.

K. Fokianos and L. Truquet. On categorical time series with covariates. *Stochastic processes and their applications*, 129:3446–3462, 2019.

Konstantinos Fokianos and Benjamin Kedem. Regression theory for categorical time series. *Statist. Sci.*, 18:357–376, 2003. ISSN 0883-4237. doi: 10.1214/ss/1076102425. URL http://dx.doi.org/10.1214/ss/1076102425.

Konstantinos Fokianos and Lionel Truquet. On categorical time series models with covariates. *Stochastic Processes and their Applications*, 2018.

Christian Francq and Le Quyen Thieu. Qml inference for volatility models with covariates. *Econometric Theory*, 35(1):37–72, 2019.

Marius Iosifescu. *Iterated function systems: A critical survey.* Univ., 2009.

Heikki Kauppi and Pentti Saikkonen. Predicting us recessions with dynamic binary response models. *The Review of Economics and Statistics*, 90(4):777–791, 2008.

Gérard Letac. A contraction principle for certain markov chains and its applications. *Contemp. Math*, 50:263–273, 1986.

Weidong Liu and Wei Biao Wu. Asymptotics of spectral density estimates. *Econometric Theory*, 26(4):1218–1245, 2010.

Weidong Liu, Wei Biao Wu, et al. Simultaneous nonparametric inference of time series. *The Annals of Statistics*, 38(4):2388–2421, 2010.

Helmut Lütkepohl. *New introduction to multiple time series analysis.* Springer Science & Business Media, 2005.

Rasmus Søndergaard Pedersen and Anders Rahbek. Testing garch-x type models. *Econometric Theory*, pages 1–36, 2018.

Jeffrey R Russell and Robert F Engle. A discrete-state continuous-time model of financial transactions prices and times: The autoregressive conditional multinomial–autoregressive conditional duration model. *Journal of Business & Economic Statistics*, 23(2):166–180, 2005.

Christopher A Sims. Money, income, and causality. *The American economic review*, 62(4):540–552, 1972.

Örjan Stenflo. A survey of average contractive iterated function systems. *Journal of Difference Equations and Applications*, 18(8):1355–1380, 2012.

Wei Biao Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154, 2005.

Wei Biao Wu. Strong invariance principles for dependent random variables. *The Annals of Probability*, 35(6):2294–2320, 2007.

Wei Biao Wu and Xiaofeng Shao. Limit theorems for iterated random functions. *Journal of Applied Probability*, 41(2):425–436, 2004.

Wei-Biao Wu and Ying Nian Wu. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics*, 10(1):352–379, 2016.

Wei Biao Wu, Yinxiao Huang, and Yibi Huang. Kernel estimation for time series: An asymptotic theory. *Stochastic Processes and their Applications*, 120(12):2412–2431, 2010.

Han Xiao and Wei Biao Wu. Covariance matrix estimation for stationary time series. *The Annals of Statistics*, 40(1):466–493, 2012.