Econometric Theory, **37**, 2021, 1135–1172.
doi:10.1017/S0266466620000559

ITERATIONS OF DEPENDENT RANDOM MAPS AND EXOGENEITY IN NONLINEAR DYNAMICS*

ZINSOU MAX DEBALY
Center for Research in Economics and Statistics
LIONEL TRUQUET
Center for Research in Economics and Statistics

We discuss the existence and uniqueness of stationary and ergodic nonlinear autoregressive processes when exogenous regressors are incorporated into 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 discuss the dependence properties of these processes using the functional dependence measure of Wu (2005, *Proceedings of the National Academy of Sciences* 102, 14150–14154) that delivers a central limit theorem giving a wide range of applications. Our results are illustrated with conditional heteroscedastic autoregressive nonlinear models, Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) processes, count time series, binary choice models, and categorical time series for which we provide many extensions of existing results.

1. INTRODUCTION

Among the various contributions devoted to time series analysis, theoretical results justifying stationarity 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 Vector AutoRegressive Moving Average (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, Hendry, and Richard (1983). For nonlinear dynamics, a few contributions consider the problem of exogenous regressors. For general GARCH-type processes, Francq and Thieu (2019) recently studied stationarity conditions when the noise and the covariate process form a stationary process. Agosto et al. (2016) considered a Poisson autoregressive process with exogenous

*This work was funded by CY Initiative of Excellence (grant "Investissements d'Avenir" ANR-16-IDEX-0008), Project "EcoDep" PSI-AAP2020-0000000013. Address correspondence to Lionel Truquet, Center for Research in Economics and Statistics, National School of Statistics and Information Analysis, UMR CNRS 9194, Campus de Ker-Lann, rue Blaise Pascal, BP 37203, 35172 Bruz Cedex, France; e-mail: Lionel.TRUQUET@ensai.Fr.

© The Author(s), 2021. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

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 on some exogenous regressors. Fokianos and Truquet (2019) studied stationarity and ergodicity of general categorical time series defined conditionally on a strictly exogenous covariate process.

In this paper, we give general results for getting stationarity, ergodicity, and stochastic dependence properties for general nonlinear dynamics defined in terms 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 model:

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

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 $f_t(x) = F\big(x, Z_{t-1}, \varepsilon_t\big)$. 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 typically 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 has been extensively studied in the independent case. In this case, the process $(X_t)_{t\in\mathbb{Z}}$ is a Markov chain. We refer 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 referred 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. We recall the following result, which can be found in Elton (1990) (see also Iosifescu, 2009, Thm. 6.2) and which generalizes a widely known result given in Brandt (1986) or Bougerol and Picard (1992) for iterations of affine random maps.

To do so, we introduce some notations and conditions. 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

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

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

Moreover, for any integers $s < t$, we set $f_s^t = f_t \circ \cdots \circ f_s$. For a positive real number x, we set $\log^+(x) = \log(x)$ if $x \geq 1$ and 0 otherwise.

THEOREM 1. *Assume that the process* $((Z_t, \varepsilon_t))_{t\in\mathbb{Z}}$ *in (1) is stationary and ergodic. Assume further that* $\mathbb{E}[\log^+ c(f_0)] < \infty$ *and* $\mathbb{E}[\log^+ d(x_0, f_0(x_0))] < \infty$ *for some point* $x_0 \in E$.

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

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

Moreover,

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

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 nonlinear random maps, Theorem 1 is less known in the time series literature. In this paper, we will make use of Theorem 1 for defining a general class of categorical time series with exogenous covariates. In particular, 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 almost surely Lipschitz. Such a property is not always valid, for instance, for the Poissonian autoregressions discussed in Section 4.3. When there are 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. Existence of some moments for the marginal distributions that are sometimes necessary for statistical applications cannot be obtained directly from this result.
3. 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, see Theorem 2 and its extension Theorem 4, is obtained by replacing the usual contraction on average condition by a contraction in conditional expectation. The assumptions that 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

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

to include exogenous regressors in nonlinear dynamics. For strictly exogenous regressors, i.e., the processes $(Z_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ are independent, we also provide an additional result, see Theorem 3, with weaker assumptions. In the context of Theorems 2 and 4, we will then 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 refer the reader to the standard textbook of Doukhan (1994, Sect. 2.4) for mixing properties of iterations of independent random functions and to Andrews (1984) for a famous counterexample of a nonstrongly mixing sequence defined via iterations of random maps. 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 and the criteria for getting mixing properties of Markov chains are useless.

This paper is mainly motivated by dynamics of type (1) with covariates that are not necessarily strictly exogenous, assuming that at any time t, the noise ε_t is independent from the past information $\sigma((Z_s,\varepsilon_s) : s \le t-1)$. The term predetermindness is sometimes used in the literature. This independence assumption is substantially weaker than the independence between the two processes $(\varepsilon_t)_{t\in\mathbb{Z}}$ and $(Z_t)_{t\in\mathbb{Z}}$. 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 on 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 $(X_s)_{s\le t}$ conditionally on $(Z_s)_{s\le t-1}$. 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 $Z_t = H(\eta_t,\eta_{t-1},\dots)$ 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 on 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, Sect. 10.2). Let us mention that there exist additional concepts of exogeneity that are introduced and discussed in Engle et al. (1983), in particular a 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 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

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

$((Z_{t-1},\varepsilon_t))_{t\in\mathbb{Z}}$ 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). Our contribution is then the first one presenting a general framework for inclusion of covariates in nonlinear dynamics. Our results can be applied to any statistical procedure, which require either ergodic properties or the use of some limit theorems developed from the notion of functional dependence introduced by Wu (2005). Since the existing literature already contains many asymptotic results of this type, we do not discuss specific applications.

The rest of this paper is organized as follows. In Section 2, we give our main results for defining stationary and ergodic solutions for recursions of type (1). In Section 3, we study weak dependence properties of the process using the functional dependence measure of Wu (2005). Many examples of nonlinear time series models satisfying our assumptions are given in Section 4, and we revisit some nonlinear dynamics discussed recently in the literature, but we also consider new ones. A conclusion is given in Section 5. The proofs of our results are postponed to the last section of the paper.

2. GENERAL RESULTS

In this section, we state several results for controlling the convergence of the backward iterations in some $\mathbb{L}^p$ spaces. We recall that for a random variable X and a real number $p \geq 1$, the quantity $\|X\|_p = \mathbb{E}^{1/p}(|X|^p)$ is called the $\mathbb{L}^p$ norm of X. Now, let $(f_t)_{t\in\mathbb{Z}}$ be 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 stochastic process taking values in another Polish space E' and $F : E \times E' \to E$ is a measurable map. In connection with our initial example (1), we have $\zeta_t = (Z_{t-1},\varepsilon_t)$. In the latter case, we will assume throughout the paper that $E' = E'_1 \times E'_2$, where E'_1 is a Borel subset of $\mathbb{R}^e$ and E'_2 is another Polish space.

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.

2.1. Conditional Contraction on Average

We first give a general and useful result for getting a convergence in some $\mathbb{L}^p$ spaces. For some real numbers $p \geq 1$, $L > 0$, and $\kappa \in (0,1)$, and an integer $m \geq 1$, we consider the two following assumptions.

A1 There exists $x_0 \in E$ such that $\sup_{t\in\mathbb{Z}} \mathbb{E}[d^p(f_t(x_0),x_0)] < \infty$.

A2 For every $t \in \mathbb{Z}$, almost surely, the following inequalities hold for every $(x,y) \in E^2$.

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

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

THEOREM 2. *Suppose that Assumptions **A1** and **A2** hold.*

1. *For every $(x,t) \in E \times \mathbb{Z}$, there exists an E-valued random variable $X_t(x)$ such that*

$$\sup_{t\in\mathbb{Z}} \|d(X_t(x),x_0)\|_p < \infty, \quad \sup_{t\in\mathbb{Z}} \|d\left(f_{t-s}^t(x),X_t(x)\right)\|_p = O\left(\kappa^{s/m}\right).$$

   Moreover, the sequence $\left(f_{t-s}^t(x)\right)_{s\geq 0}$ converges almost surely to $X_t(x)$.
2. *For $x \neq y$, we have $\mathbb{P}(X_t(x) \neq X_t(y)) = 0$. We then set $X_t = X_t(x)$.*
3. *The process $((X_t,\zeta_t))_{t\in\mathbb{Z}}$ is stationary and also ergodic if the process $(\zeta_t)_{t\in\mathbb{Z}}$ is itself stationary and ergodic.*
4. *If $(Y_t)_{t\in\mathbb{Z}}$ is a nonanticipative process (i.e., $Y_t \in \mathcal{F}_t$) such that $Y_t = f_t(Y_{t-1})$ for $t \in \mathbb{Z}$ and $\sup_{j\in\mathbb{Z}} \mathbb{E}\left[d^p\left(Y_j,x_0\right)\right] < \infty$, then $Y_t = X_t$ a.s.*

Notes

1. The bounds given in Assumption **A2** are required to hold for all $(x,y) \in E^2$ at the same time. Since a conditional expectation is only unique up to a set with measure 0 for the probability measure $\mathbb{P}$, the bound given in Assumption **A2** has to be understood in terms of regular conditional distribution, i.e., there exists a regular version of the conditional distribution of $(\zeta_t,\ldots,\zeta_{t+m-1})$ given $\mathcal{F}_{t-1}$. On Polish spaces, a regular version always exists. See Kallenberg (2006, Chap. 5).
2. When $\zeta_t = (Z_{t-1},\varepsilon_t)$ forms a stationary process and $\mathcal{F}_t = \sigma\left((Z_j,\varepsilon_j) : j \leq t\right)$, Theorem 2 guarantees existence and uniqueness of a stationary process possessing a moment of order p and solution of (1). However, stationarity of the covariate/error process is not required for applying this result. In particular, when $(\zeta_t)_{t\in\mathbb{Z}}$ is nonstationary, one can still define solutions of (1) provided that **A1** and **A2** are satisfied. In general, these solutions will be nonstationary, and the classical law of large numbers is not valid. In this case, studying the asymptotic properties of some classical inferential procedures, such as conditional likelihood estimation, requires a specific analysis.
3. Setting $p = 1$ and $d(x,y) = |x-y|^o$ for $x,y \in E = \mathbb{R}$ and some $o \in (0,1)$, one can consider stochastic recursions (1) with heavy-tailed covariate processes.
4. When the process $(\zeta_t)_{t\in\mathbb{Z}}$ is stationary and the recursions are initialized at time $t = 0$ with a given state $x \in E$, the probability distribution of the forward iterations $f_1^t(x)$ coincides with the probability distribution of the backward iterations $f_{-t+1}^0(x)$. Since $\lim_{t\to\infty} f_{-t+1}^0(x) = X_0$ a.s., $f_1^t(x)$ converges in distribution to X_0. The same property holds true when the iterations are initialized with a random variable $\overline{X}_0$ independent from $(\zeta_t)_{t\geq 1}$. The main interest of the convergence of the backward iterations is to define the good random initialization $X_0 = f_{-\infty}^0(x)$ in order to get a stationary process $(X_t)_{t\geq 0}$.
5. Assumption **A2** is a conditional contraction property in $\mathbb{L}^p$ which is crucial for getting the convergence of the backward iterations in $\mathbb{L}^p$ norms. Relaxing this assumption by introducing a random coefficient κ_{t-1} instead of κ can be problematic for getting a similar result. We discuss this point below.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

2.2. Comments on Assumption A2

Let us consider the dynamic (1), set $f_t(x) = F\big(x, Z_{t-1}, \varepsilon_t\big)$, with $E = \mathbb{R}$, and $p = 1$, and assume that for every t, ε_t is independent from $\mathcal{F}_{t-1} = \sigma\left((Z_j, \varepsilon_j) : j \le t-1\right)$. Then, Assumption **A2** is satisfied for $m = 1$ if and only if there exists $\kappa \in (0,1)$ such that

$$\sup_{z \in E_1'} \mathbb{E}\,|F(x,z,\varepsilon_0) - F(y,z,\varepsilon_0)| \le \kappa |x-y|.$$

At a first sight, the latter condition is quite strong, and it is natural to wonder if the following weaker assumption can be used, i.e., there exists a measurable function $\kappa : G \to (0,\infty)$ such that for every $z \in G$,

$$\mathbb{E}\,|F(x,z,\varepsilon_0) - F(y,z,\varepsilon_0)| \le \kappa(z)|x-y|.$$

Of course, the challenging question concerns the convergence of the backward iterations when the function κ may take values larger than 1. However, a problem occurs for applying the successive contraction properties to the iterated random maps. Consider the iterations $f_t \circ f_{t-1}$. We have for $(x,y) \in E^2$,

$$\begin{aligned}\mathbb{E}\,|f_t \circ f_{t-1}(x) - f_t \circ f_{t-1}(y)| &= \mathbb{E}\left[\mathbb{E}\left[|f_t \circ f_{t-1}(x) - f_t \circ f_{t-1}(y)|\,\mathcal{F}_{t-1}\right]\right] \\ &\le \mathbb{E}\left[\kappa(Z_{t-1})\,|f_{t-1}(x) - f_{t-1}(y)|\right].\end{aligned}$$

If the random variable $\kappa(Z_{t-1})$ depends on past values of the error $\varepsilon_{t-j}, j \ge 1$, it is stochastically dependent on the random map f_{t-1} and also on $\mathcal{F}_{t-2}$. It is then not possible to use the contraction property of f_{t-1} unless the function κ can be bounded by a constant. To show that the successive iterations lose the memory with respect to initialization, this constant has to be smaller than 1. Of course, this does not prove that the convergence of the iterations in $\mathbb{L}^1$ is not possible.

To show that the convergence of the backward iterations in $\mathbb{L}^1$ is problematic, we now consider a map f_t linear in x, a case for which an explicit solution is available. Accordingly, we assume that

$$f_t(x) = \kappa(Z_{t-1})x + \varepsilon_t,$$

where the function κ is bounded but not necessarily by 1 and an integrable noise ε_0. The dynamic is then given by an AutoRegressive (AR) process with a random lag coefficient, and it is widely known that the unique solution can be written as

$$X_t = \sum_{j \ge 1} \prod_{i=1}^{j} \kappa(Z_{t-i})\,\varepsilon_{t-j} + \varepsilon_t, \tag{2}$$

provided that $\mathbb{E}\log\kappa(Z_0) < 0$. The series (2) converges almost surely. The $\mathbb{L}^1$ convergence is guaranteed from Theorem 2, as soon as $\kappa := \|\kappa(Z_0)\|_\infty < 1$, where for a random variable X, $\|X\|_\infty$ denotes its supremum norm. If $\kappa \ge 1$, convergence in $\mathbb{L}^1$ of the series 2 is much more difficult to get because of the possible stochastic dependence between the coordinates of the process $(Z_t)_{t \in \mathbb{Z}}$. Let us first note that

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

such a problem occurs in the nonergodic case, when $Z_t = Z_0$ a.s. In this case, we have

$$\mathbb{E}\left(\prod_{i=1}^{j}\kappa\left(Z_{t-i}\right)|\varepsilon_{t-j}|\right) = \mathbb{E}\left(|\varepsilon_0|\right)\cdot\mathbb{E}\left(\kappa(Z_0)^j\right)$$

and since $\mathbb{E}\left(\kappa(Z_0)^j\right) \geq \mathbb{P}\left(\kappa(Z_0) \geq 1\right)$, one cannot get convergence of the series (2) if $\mathbb{P}\left(\kappa(Z_0) \geq 1\right) > 0$. In what follows, we also stress that a similar problem of convergence also occurs in the ergodic case. To this end, set $\phi(p) = \|\kappa(Z_0)\|_p$ for $p \geq 1$. The function ϕ is nondecreasing and $\phi(\infty) = \|\kappa(Z_0)\|_\infty$. Assumption **A2** is satisfied as soon as $\|\kappa(Z_0)\|_\infty < 1$. It is then tempting to study the convergence of the solution only assuming that $\|\kappa(Z_0)\|_p < 1$ but $\|\kappa(Z_0)\|_q \geq 1$ for some $1 \leq p < q$. However, for any value of the pair (p,q), there always exists an example of a process $(\kappa(Z_t))_{t\in\mathbb{Z}}$ such that the series (2) is not converging in $\mathbb{L}^1$. To this end, assuming without loss of generality that q is an integer, we define $\kappa(z) = z$ and $Z_{t-1} = a_{t-1}\cdots a_{t-q}$, where $(a_t)_{t\in\mathbb{Z}}$ is a process of i.i.d. nonnegative random variables, independent of $(\varepsilon_t)_{t\in\mathbb{Z}}$, such that $\|a_0\|_p < 1$ and $\|a_0\|_q \geq 1$. Since for $j \geq q$,

$$\prod_{i=1}^{j}\kappa\left(Z_{t-i}\right) = \prod_{i=1}^{q-1}a_{t-i}^{i}\cdot\prod_{i=q}^{j+1}a_{t-i}^{q}\prod_{i=0}^{q-2}a_{t-j-k+i}^{i+1},$$

we find

$$\mathbb{E}\left(\prod_{i=1}^{j}\kappa\left(Z_{t-i}\right)\right) = \prod_{i=1}^{q-1}\mathbb{E}^2\left(a_0^i\right)\cdot\mathbb{E}^{j-q+2}\left(a_0^q\right).$$

Hence, the previous expectation does not converge to 0 when $j \to \infty$, and the series (2) cannot converge in $\mathbb{L}^1$. The analysis of this linear case enlightens that a tail condition on $\kappa(Z_t)$ is not sufficient for getting this kind of convergence. In particular, the dependence structure of the process $(\kappa(Z_t))_{t\in\mathbb{Z}}$ is also of major importance. This contrasts with AR processes with i.i.d. random coefficients, since in this case the condition $\mathbb{E}\kappa(Z_0) < 1$ is necessary and sufficient for the convergence of the series (2) in $\mathbb{L}^1$. However, imposing an independence assumption on the covariate process is not reasonable.

In the next section, we show that one can investigate a different mode of convergence for the iterations in model (1) and which allow to relax Assumption **A2**. However, it is necessary to impose a strict exogeneity assumption on the covariate process.

2.3. An Additional Result for Strictly Exogenous Regressors

In this subsection, we consider specifically equation (1) when the covariate process is independent of the error process. In this case, conditionally on Z, the process is a time-inhomogeneous Markov chain. The terminology Markov chain in random

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

environments is often used in the literature. See for instance Stenflo (2001). The following result will not be central in the rest of the paper, because substantial efforts could be needed to derive moment and weak dependence properties for the corresponding solution and it could be also difficult to obtain explicit conditions for dealing with higher-order autoregressive processes. This is why we only provide a result when the f_t's satisfied a one-step contraction (i.e., $m = 1$ in **A2**). We assume that there exist a real number $p \geq 1$ and a state $x_0 \in E$ such that the three following conditions are fulfilled.

A0 The process $Z := (Z_t)_{t\in\mathbb{Z}}$ is stationary and ergodic, and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a process of i.i.d. random variables taking values in E_2' and is independent of Z.

A1' For every $z \in E_1'$, we have $\mathbb{E}[d^p(F(x_0, z, \varepsilon_1), x_0)] < \infty$.

A2' There exists a measurable function $\kappa : E' \to (0, \infty)$ satisfying $\mathbb{E}(\log^+ \kappa(Z_0)) < \infty$, $\mathbb{E}\log\kappa(Z_0) < 0$ such that for every $(x, y) \in E^2$,

$$\mathbb{E}\left[d^p(F(x, z, \varepsilon_0), F(y, z, \varepsilon_0))\right] \leq \kappa^p(z) d^p(x, y).$$

Moreover,

$$\mathbb{E}\left[\log^+ \int d^p\big(x_0, F(x_0, Z_0, u)\big) d\mathbb{P}_{\epsilon_1}(u)\right] < \infty.$$

We remind the notation $f_t(x) = F\big(x, Z_{t-1}, \varepsilon_t\big)$. Here, we set $\mathcal{F}_t = \sigma\left((Z_j, \varepsilon_j) : j \leq t\right)$ and $\mathbb{E}[X|Z]$ will denote the expectation of a random variable X conditionally on the covariate process $(Z_t)_{t\in\mathbb{Z}}$.

THEOREM 3. *Suppose that Assumptions **A0, A1', and A2'** hold.*

1. *For every $(x,t) \in E \times \mathbb{Z}$, there exists a random variable $X_t(x)$ such that $\mathbb{E}[d^p(X_t(x), x_0)|Z] < \infty$ a.s. and $\lim_{s\to\infty} \mathbb{E}\left[d^p\left(f_{t-s}^t(x), X_t(x)\right)|Z\right] = 0$ a.s. The sequence $\left(f_{t-s}^t(x)\right)_{s\geq 0}$ also converges almost surely to $X_t(x)$.*
2. *For $x \neq y$, we have $\mathbb{P}(X_t(x) \neq X_t(y)) = 0$. We then set $X_t = X_t(x)$.*
3. *The process $((X_t, Z_t))_{t\in\mathbb{Z}}$ is stationary and ergodic.*
4. *If $(Y_t)_{t\in\mathbb{Z}}$ is a nonanticipative process (i.e., $Y_t \in \mathcal{F}_t$) such that $((Y_t, Z_t))_{t\in\mathbb{Z}}$ is stationary, for every $t \in \mathbb{Z}$, $Y_t = f_t(Y_{t-1})$ and $\mathbb{E}[d^p(Y_0, x_0)|Z_0, Z_{-1}, \ldots] < \infty$ a.s., then $Y_t = X_t$ a.s.*

Notes

1. The contraction inequality in Assumption **A2'** can be restated as

   $$\mathbb{E}\left[d^p(f_t(x), f_t(y))|Z\right] \leq \kappa^p(Z_{t-1}) d^p(x, y) \text{ a.s.}$$

   It is then another example of contraction in conditional average.
2. In our context, our result can be seen as an improvement of Thm. 1 given in Stenflo (2001) for Markov chains in random environments. In particular, we do not assume a uniform contraction with respect to the environment, which is given by the exogenous process Z in our random maps f_t.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

2.4. Example

We compare the contraction conditions necessary to apply Theorems 1—3 on a specific example. Let $((\varepsilon_t, Z_t))_{t\in\mathbb{Z}}$ be a stationary sequence of pair of random variables and for $t \in \mathbb{Z}$, set $\mathcal{F}_t = \sigma\left((Z_j, \varepsilon_j) : j \le t\right)$. Assume that $\mathbb{E}\left(\varepsilon_1^2|\mathcal{F}_0\right) = 1$ and $\mathbb{E}\left(\varepsilon_1|\mathcal{F}_0\right) = 0$. For $i = 1, 2$, let $a_i : E_1' \to \mathbb{R}$ and $b_i : E_1' \to \mathbb{R}_+$ be some measurable maps such that $\mathbb{E}\log^+ a_i(Z_0) < \infty$ and $\mathbb{E}\log^+ b_i(Z_0) < \infty$. We consider the following AR-ARCH model with functional coefficients:

$$X_t = a_0(Z_{t-1}) + a_1(Z_{t-1})X_{t-1} + \varepsilon_t\sqrt{b_0(Z_{t-1}) + b_1(Z_{t-1})X_{t-1}^2}.$$

Here, we set $E = \mathbb{R}$, $d(x, y) = |x - y|$, and $p = 2$. Setting $\sigma(z, x) = \sqrt{b_0(z) + b_1(z)x^2}$, one can note that $|\sigma(z, x) - \sigma(z, y)| \le \sqrt{b_1(z)}|x - y|$.

1. To apply Theorem 1, we compute the Lipschitz constant $c(f_1)$ of the random map f_1. We have

$$c(f_1) = \sup_{v\in\mathbb{R}} |f_1'(v)| = \sup_{v\in\mathbb{R}} \left| a_1(Z_0) + \frac{\varepsilon_1 b_1(Z_0)v}{\sqrt{b_0(Z_0) + b_1(Z_0)v^2}} \right|.$$

   Making the change of variable $\overline{v} = sign(\varepsilon_1)sign\left(a_1(Z_0)\right)v$, we have

$$c(f_1) = \sup_{\overline{v}\in\mathbb{R}} \left[|a_1(Z_0)| + \frac{|\varepsilon_1| b_1(Z_0)\overline{v}}{\sqrt{b_0(Z_0) + b_1(Z_0)\overline{v}^2}} \right|.$$

   We then obtain $c(f_1) = |a_1(Z_0)| + \sqrt{b_1(Z_0)}|\varepsilon_1|$, and Theorem 1 applies as soon as

$$\mathbb{E}\log c(f_1) < 0. \tag{3}$$

2. To apply Theorem 2, note that

$$\begin{aligned}\mathbb{E}\left[|f_t(x) - f_t(y)|^2|\mathcal{F}_{t-1}\right] &= a_1(Z_{t-1})^2(x - y)^2 + (\sigma(Z_{t-1}, x) - \sigma(Z_{t-1}, x))^2 \\ &\le \left(a_1(Z_{t-1})^2 + b_1(Z_{t-1})\right) \cdot |x - y|^2.\end{aligned}$$

   One can then show that Theorem 2 applies as soon as

$$\kappa^2 := \sup_z \left(a_1(z)^2 + b_1(z)\right) < 1. \tag{4}$$

3. If the two processes $(\varepsilon_t)_{t\in\mathbb{Z}}$ and $(Z_t)_{t\in\mathbb{Z}}$ are independent and the ε_t's are i.i.d., Theorem 3 applies as soon as

$$\mathbb{E}\log\left(a_1(Z_0)^2 + b_1(Z_0)\right) < 0. \tag{5}$$

Under the strict exogeneity assumption, we note that (5) is weaker than (4). However, (4) ensures the existence of a second-order moment for the solution, whereas (5) only guarantees that $\mathbb{E}\left(X_t^2|Z\right) < \infty$ a.s. On the other hand, (3), which only ensures existence of a stationary solution, is not necessarily weaker than (4) or (5). For instance, if the noise process has a Rademacher distribution,

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

$\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = 1/2$ and the functional coefficients are constant, (3) writes $|a_1| + \sqrt{b_1} < 1$, which is more restrictive than (4) or (5). But if a_1 is identically equal to 0, (3) writes as $\frac{1}{2}\mathbb{E}\log b_1(Z_0) + \mathbb{E}\log|\varepsilon_0| < 0$, which is weaker than (5), since from Jensen's inequality, we have

$$\mathbb{E}\log|\varepsilon_0| \le \log\mathbb{E}|\varepsilon_0| \le \log\mathbb{E}^{1/2}\left(\varepsilon_0^2\right) = 0.$$

2.5. A Result for Higher-Order Autoregressions

In this subsection, we extend Theorem 2 to higher-order autoregressive processes. We only consider stationary processes in this part. The main result, Theorem 4, is particularly interesting for multivariate autoregressions for which Lipschitz-type properties can be obtained equation by equation. See Section 4 for an application of Theorem 4 to various examples.

For a given real number $0 < o \le 1$, we define the distance Δ on $\mathbb{R}$, by $\Delta(u,v) = |u-v|^o$ for $u, v \in \mathbb{R}$. Let E be a subset of $\mathbb{R}^k$ and $\|\cdot\|$ an arbitrary norm on $\mathbb{R}^k$. Our aim is to study existence of solutions for the following recursive equations:

$$X_t = F\left(X_{t-1}, \ldots, X_{t-q}, \zeta_t\right), \quad t \in \mathbb{Z}, \tag{6}$$

where $F : E^q \times E' \to E$ is a measurable function. Note that one can always associate a random map f_t on E^q to the dynamic (6). To this end, for $t \in \mathbb{Z}$ and $x_1, \ldots, x_q \in E$, we set

$$f_t(x_1, \ldots, x_q) = \left(F\left(x_1, \ldots, x_q, \zeta_t\right), x_1, \ldots, x_{q-1}\right).$$

We first introduce additional notations. Let $\mathcal{M}_k$ be the set of square real matrices of order k rows and if $A \in \mathcal{M}_k$, $\rho(A)$ the spectral radius of the matrix A. Moreover, for $x, y \in \mathbb{R}^k$ and $p \ge 1$, the vector $(\Delta^p(x_1, y_1), \ldots, \Delta^p(x_k, y_k))'$ will be denoted by $\Delta^p_{vec}(x, y)$. Finally, we introduce a partial order relation $\preceq$ on $\mathbb{R}^k$ such that $x \preceq y$ means $x_i \le y_i$ for $i = 1, \ldots, k$.

The following assumptions will be needed.

B1 The process $(\zeta_t)_{t\in\mathbb{Z}}$ is stationary and ergodic adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$.
B2 For any $y \in E^q$, $\mathbb{E}[\|F(y, \zeta_1)\|^{op}] < \infty$.
B3 There exist some matrices $A_1, \ldots, A_q \in \mathcal{M}_k$ with nonnegative elements, satisfying $\rho\left(A_1 + \cdot + A_q\right) < 1$, such that for $y, y' \in E^q$,

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

Though the following result is stated for an arbitrary pair $o \in (0,1), p \ge 1$, the two interesting cases are $o \in (0,1), p = 1$ and $o = 1, p > 1$.

THEOREM 4. *Suppose that Assumptions **B1–B3** hold. There then exists a unique stationary and nonanticipative process $(X_t)_{t\in\mathbb{Z}}$ solution of (6) such that $\mathbb{E}[\|X_t\|^{op}] < \infty$. Moreover, this process is ergodic.*

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

Notes

1. If $\zeta_t = (Z_{t-1}, \varepsilon_t)$ with ε_t independent of $\mathcal{F}_{t-1} = \sigma\left((Z_j, \varepsilon_j) : j \le t\right)$ and $k = o = 1$, Assumption **B3** writes

$$\sup_z \mathbb{E}\left[\left|F(y_1, \ldots, y_q, z, \varepsilon_1) - F(y_1', \ldots, y_q', z, \varepsilon_1)\right|^p\right] \le \sum_{i=1}^{q} A_i |y_i - y_i'|^p,$$

   with $\rho\left(\sum_{i=1}^q A_i\right) = \sum_{i=1}^q A_i < 1$. This provides a quite simple criterion for application to autoregressive processes.
2. In the spirit of Section 2.4, the previous criterion can be checked for models with varying parameters, directly constructed from smooth parametric autoregressive processes. Consider the model

$$Y_t = \overline{F}_\theta\left(Y_{t-1}, \ldots, Y_{t-q}, \varepsilon_t\right), \quad t \in \mathbb{Z}, \theta \in \Theta \subset \mathbb{R}^{\bar{e}}.$$

   If

$$\mathbb{E}\left[\left|\overline{F}_\theta(y_1, \ldots, y_q, \varepsilon_1) - \overline{F}_\theta(y_1', \ldots, y_q', \varepsilon_1)\right|^p\right] \le \sum_{i=1}^{q} \overline{A}_i(\theta) |y_i - y_i'|^p,$$

   the model

$$Y_t = \overline{F}_{\theta(Z_{t-1})}\left(Y_{t-1}, \ldots, Y_{t-q}, \varepsilon_t\right)$$

   satisfies **B3** as soon as $\sum_{i=1}^q \|\overline{A}_i\|_\infty < 1$. We then obtain a model with exogenous covariates by replacing parameter θ with a varying parameter $\theta(Z_{t-1})$, where $\theta : \mathbb{R}^e \to \Theta$ is a measurable map. See also the note after Proposition 3 for a discussion.

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., $X_t = H(\xi_t, \xi_{t-1}, \ldots)$, where $(\xi_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables taking values in a measurable space $(G, \mathcal{G})$. The functional dependence measure is expressed in terms of some coefficients which evaluate for $t \ge 0$ the $\mathbb{L}^p$ distance between X_t and a copy $\overline{X}_t$, obtained by replacing ξ_0 with ξ_0', ξ_0' following the same distribution as ξ_0 and being independent from the sequence $(\xi_t)_{t\in\mathbb{Z}}$. See below for the definition of these coefficients. Most of the limit theorems and deviation inequalities have been derived under such dependence measures. See for instance Wu (2007) and Wu and Wu (2016). Such asymptotic results have been applied to various statistical problems. See for instance Wu, Huang, and Huang (2010) for kernel estimation for time series, Xiao and Wu (2012) for covariance estimation,

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

or Liu and Wu (2010a) 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 section is to show that under the assumptions of Theorem 2 or 4, when the process $(\zeta_t)_{t\in\mathbb{Z}}$ or more specifically the covariate process $(Z_t)_{t\in\mathbb{Z}}$ satisfies this kind of dependence, the functional dependence measure of the solution $(X_t)_{t\in\mathbb{Z}}$ can be controlled. 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 the state space E is a subspace of $\mathbb{R}^k$ and the distance d is given by $d(x,y) = |x-y|^o$, where $|\cdot|$ is an arbitrary norm on $\mathbb{R}^k$ and $0 < o \le 1$.

3.1. Dependence Coefficients for General Iterated Random Functions

Assume that the process ζ has a Bernoulli shift representation, i.e., $\zeta_t = H(\xi_t, \xi_{t-1}, \ldots)$ for some measurable map $H : G^{\mathbb{N}} \to E'$ and $(\xi_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d., G-valued random variables. We then have for every $(x,t) \in E \times \mathbb{Z}$,

$$f_t(x) = F(x, H(\xi_t, \xi_{t-1}, \ldots))$$

and the map f_t has itself a Bernoulli shift representation. To define the functional measure coefficients, we then define a new sequence $\left(\overline{\xi}_t\right)_{t\in\mathbb{Z}}$ such that $\overline{\xi}_0 = \xi_0'$ and $\overline{\xi}_t = \xi_t$ for $t \neq 0$. Here, ξ_0' is a copy of ξ_0, which is assumed to be independent from $(\xi_t)_{t\in\mathbb{Z}}$. Moreover, for $t > 0$, let

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

We define for $t \ge 0$ and $p \ge 1$,

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

Moreover, for $h \in \mathbb{N}$, let $\Theta_{p,h} = \sum_{t\ge h}\theta_{p,t}$. Two cases of interest are $p = 1$ and $o = 1, p > 1$.

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

A3 There exists a measurable function $S : E \to \mathbb{R}_+$ and $r, s \ge p$ 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 $t \ge 1$,

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

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

An immediate consequence of Assumption **A4** is that for any random variable V_{t-1}, measurable with respect to $\sigma(\xi_0') \vee \mathcal{F}_{t-1}$, we have

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

$\mathbb{E}\left[d^p\left(\overline{f}_t(V_{t-1}),f_t(V_{t-1})\right)\right] \le \mathbb{E}\left[S(V_{t-1})^p H_{t-1}^p\right]$ and from Hölder's inequality, we get

$$\|d\left(\overline{f}_t(V_{t-1}),f_t(V_{t-1})\right)\|_p \le \|S(V_{t-1})\|_s \|H_{t-1}\|_r.$$

When $\zeta_t = (Z_{t-1},\varepsilon_t)$, with Z_t taking values in a Borel subset E_1' of $\mathbb{R}^e$ and ε_t taking values in a Polish space E_2', we will still denote by $|\cdot|$ an arbitrary norm on $\mathbb{R}^e$ and we also set $d(z,z') = |z-z'|^o$ for $z,z' \in \mathbb{R}^e$ to avoid additional notations. We will use two specific assumptions.

A3' There exists a measurable function $S : E \to \mathbb{R}_+$ and $r,s \ge p$ such that $r^{-1} + s^{-1} = p^{-1}$ and $\mathbb{E}\left[S(f_{-\infty}^0)^s\right] < \infty$, $\mathbb{E}|Z_0|^{ro} < \infty$ and for all $x \in E$ and $t \ge 1$,

$$\mathbb{E}^{1/p}\left[d^p\left(F\left(x,z,\varepsilon_0\right),F\left(x,z',\varepsilon_0\right)\right)\right] \le S(x)d(z,z'),$$

A4 Let $(\eta_t)_{t\in\mathbb{Z}}$ be a sequence of random variables taking values in a measurable space $(G_1,\mathcal{G}_1)$ such that $Z_t = H'\left(\eta_t,\eta_{t-1},\ldots\right)$ 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. of random variables taking values in $G = E_2' \times G_1$.

From **A4**, we have the specific representation

$$\zeta_t = H\left(\xi_t,\xi_{t-1},\ldots\right) := \left(H'\left(\eta_{t-1},\eta_{t-2},\ldots\right),\varepsilon_t\right), \quad t \in \mathbb{Z}.$$

The map H takes values in $E' = E_1' \times E_2'$. Note that our formulation allows the covariate process to have a general form, including a VARMA or GARCH process among others.

PROPOSITION 1.

1. *Suppose that Assumptions **A1–A4** hold. For any $h \ge 2$, there then exists $C_1 > 0$ not depending on h such that*

$$\Theta_{p,h} \le C_1\left[\kappa^{h/m} + \sum_{i=0}^{h-1}\kappa^{i/m}\eta_{r,h-i} + \sum_{i\ge h}\kappa^{i/m}\eta_{r,1}\right], \tag{7}$$

with

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

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

$$\Theta_{p,h} \le C\left[\kappa^{h/m} + \sum_{i=0}^{h-1}\kappa^{i/m}\eta_{r,h-i}\right]. \tag{8}$$

2. *Suppose that Assumptions **A1, A2, A4,** and **A3'** hold. Then, we get the bound (8) with $\eta_{r,j} = \Theta_{r,j-1}(Z)$.*

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

Notes

1. Let us comment on Assumption **A4**. 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 simultaneous dependence between ε_t and η_t. For instance, we can set $\eta_t = K(\varepsilon_t, U_t)$, where K is a measurable function and U is a sequence of i.i.d. random variables, independent from the sequence ε. This assumption is then more flexible than the complete independence between the two error processes ε and η, which implies strict exogeneity.
2. It can happen that our assumptions are satisfied with some p, leading to an upper bound for the functional dependence coefficients $\theta_{p,t}$, 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 one can prove finiteness of higher-order moments for the solution, e.g., if $\mathbb{E}[|f_{-\infty}^0|^{q'o}] < \infty$ for some $q' > q$. Indeed, from Hölder's inequality, we have

$$\theta_{q,t} \leq \theta_{p,t}^{\frac{p(q'-q)}{q(q'-p)}} \theta_{q',t}^{\frac{q'(q-p)}{q(q'-p)}}.$$

Moreover, $\theta_{q',t} \leq 2\| \left|f_{-\infty}^0\right|^o \|_{q'}$.

3.2. Dependence Coefficients for Higher-Order Autoregressions with Exogenous Covariates

Here, we revert to higher-order autoregressions considered in Section 2.5. We consider directly the case $\zeta_t = (Z_{t-1}, \varepsilon_t)$ with Assumption **A4** being satisfied. In addition, to Assumptions **B1, B2, B3,** and **A4**, the following assumption, which is the analog of **A3'**, will be needed.

B4 If $r > 0$ and $s \in \mathbb{R}_+ \cup \{\infty\}$ such that $r^{-1} + s^{-1} = p^{-1}$, there exists a measurable function $S : E^q \to \mathbb{R}$ such that $\mathbb{E}\left[S(X_q, \ldots, X_1)^s\right] < \infty$, $\mathbb{E}|Z_0|^{or} < \infty$ and for all $z, z' \in E_1'$ and $x_1, \ldots, x_q \in E^q$,

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

The result is the following.

PROPOSITION 2. *Suppose that Assumptions* **B2**–**B4** *and* **A4** *hold. There then exists* $C > 0$ *and* $\rho \in (0,1)$*, such that for all* $h \geq 1$, $\Theta_{p,h}(X) \leq C[\rho^h + \sum_{i=1}^h \rho^i \Theta_{r,h-i}(Z)]$.

Note

From the upper bound given in Proposition 2, we note that the decay of $\Theta_{p,h}(X)$ is polynomial (respectively geometric) in h if the decay of $\Theta_{r,h}$ is polynomial (respectively geometric) in h.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

3.3. A Central Limit Theorem

To illustrate the usefulness of our results, we give below a central limit theorem for partial sums

$$S_n = \sum_{i=1}^{n} f\left(X_t, X_{t-1}, \ldots, X_{t-k}\right),$$

where f is some real-valued measurable function and $(X_t)_{t\in\mathbb{Z}}$ is a stochastic process solution of $X_t = F\left(X_{t-1}, \ldots, X_{t-q}, \zeta_t\right)$ and the assumptions of either Proposition 1 or Proposition 2 are satisfied. The following result, which is a straightforward corollary of the invariance principle given in Wu (2005), is not the most general as possible. In particular, when a moment of order greater than p is available for the stationary solution, different assumptions on the function f could be used.

THEOREM 5. *Suppose that either Assumptions **A1–A4** or Assumptions **B2–B4** and **A4** hold true for some $p > 2$ and $\Theta_{p,0}(X) < \infty$. If there exist $C > 0$ and $0 \le \ell \le \frac{p-2}{2}$ such that for $x_i, x_i' \in E$, $0 \le i \le k$,*

$$\left|f\left(x_0, \ldots, x_k\right) - f\left(x_0', \ldots, x_k'\right)\right| \le C\left[1 + \sum_{i=0}^{k}\left(|x_i|^{o\ell} + |x_i'|^{o\ell}\right)\right] \cdot \sum_{i=0}^{k} d(x_i, x_i').$$

Then, we have the weak convergence

$$\frac{1}{\sqrt{n}}\left(S_n - \mathbb{E}S_n\right) \Rightarrow \mathcal{N}\left(0, \sigma^2\right),$$

with $\sigma^2 = \sum_{j\in\mathbb{Z}} Cov\left(Y_j, Y_0\right)$, $Y_t = f\left(X_t, \ldots, X_{t-k}\right)$.

4. EXAMPLES

4.1. CHARN Models

In this section, we consider conditional heteroscedastic autoregressive nonlinear (CHARN) models, such as in Härdle and Tsybakov (1997) or Härdle, Tsybakov, and Yang (1998), but that can encompass exogenous regressors. More precisely, we consider the dynamic

$$Y_t = f_1\left(Y_{t-1}, \ldots, Y_{t-q}, Z_{t-1}\right) + \varepsilon_t f_2\left(Y_{t-1}, \ldots, Y_{t-q}, Z_{t-1}\right), \tag{9}$$

where q is a positive integer and $f_1, f_2 : \mathbb{R}^q \times E_1' \to \mathbb{R}$ are measurable functions. In order to study stationary solutions of the recursions (9), the following assumptions will be needed.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

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

CH2 For $j = 1, 2$, there exist measurable functions $a_{i,j} : \mathbb{R}^d \times \mathbb{R}_+$, $1 \le i \le q$ such that

$$\left|f_j(y_1, \ldots, y_q, z) - f_j(y'_1, \ldots, y'_q, z)\right| \le \sum_{i=1}^{q} a_{i,j}(z)\left|y_i - y'_i\right|.$$

CH3 There exist a real number $p \ge 1$ such that $\|\varepsilon_1\|_p < \infty$ and $r, s \ge p$ such that $r^{-1} + s^{-1} = p^{-1}$, s can be infinite, $\mathbb{E}|Z_0|^r < \infty$ and two functions L_1, L_2 defined on $\mathbb{R}^q$ such that for $j = 1, 2$, $y_1, \ldots, y_q \in \mathbb{R}$ and $z, z' \in E'_1$,

$$\left|f_j(y_1, \ldots, y_q, z) - f_j(y_1, \ldots, y_q, z')\right| \le L_j(y_1, \ldots, y_q)|z - z'|.$$

For $t \in \mathbb{Z}$ and $i = 1, \ldots, q$, we set $c_{i,t} = a_{i,1}(Z_{t-1}) + a_{i,2}(Z_{t-1})|\varepsilon_t|$, 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 an arbitrary norm on the space of square matrices of size $q \times q$.

PROPOSITION 3. *Suppose that Assumptions **CH1** and **CH2** hold.*

1. *Suppose that $\chi(\mathbf{A}) < 0$. There then exists a unique stationary process $(Y_t)_{t\in\mathbb{Z}}$ solution of (9), which is also ergodic.*
2. *Assume additionally that for every $t \in \mathbb{Z}$, ε_t is independent from $\mathcal{F}_{t-1}$. If there exist $\bar{x} \in \mathbb{R}^q$, $p \ge 1$ such that $f_1(\bar{x}, Z_0) + \varepsilon_1 f_2(\bar{x}, Z_0) \in \mathbb{L}^p$ and*

$$\sum_{i=1}^{q} \sup_z \|a_{i,1}(z) + a_{i,2}(z)|\varepsilon_1|\|_p < 1, \tag{10}$$

   there then exists a unique stationary and nonanticipative process solution of (9), which is also ergodic such that $\mathbb{E}|Y_1|^p < \infty$.
3. *Assume furthermore that Assumptions **CH3** and **A4** hold true with $\mathbb{E}L_j^s(Y_q, \ldots, Y_1) < \infty$ for $j = 1, 2$. There then exists $C > 0$ and $\rho \in (0, 1)$, such that for all $h \ge 1$, $\Theta_{p,h}(X) \le C\left[\rho^h + \sum_{i=1}^{h} \rho^i \Theta_{r,h-i}(Z)\right]$.*

Notes

1. Our results can be useful for dealing with models with functional coefficients in the spirit of the example given in Section 2.4. See also the notes after the

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

statement of Theorem 4. For $j = 1, 2$, let $m_\theta^{(j)} : \mathbb{R}^q \to \mathbb{R}$ be some functions depending on some parameters $\theta \in \mathbb{R}^e$ such that

$$\left| m_\theta^{(j)}(y_1, \ldots, y_q) - m_\theta^{(j)}(y_1', \ldots, y_q') \right| \leq \sum_{i=1}^{q} d_{i,j}(\theta) \left| y_i - y_i' \right|,$$

for some nonnegative real numbers $d_{i,j}(\theta)$, $1 \leq i \leq q$. If θ is replaced by a function $\theta(\cdot) : \mathbb{R}^d \to \mathbb{R}^e$ and $f_j(y_1, \ldots, y_q, z) = m_{\theta(z)}^{(j)}(y_1, \ldots, y_q)$ for $j = 1, 2$, one can then consider some standard autoregressive processes and obtain a version with functional parameters depending on exogenous covariates. For instance, threshold autoregressions or power AutoRegressive conditional Heteroskedasticity (ARCH) volatility,

$$m_\theta^{(1)}(y_1, \ldots, y_q) = \theta_0 + \sum_{i=1}^{q} \left(\theta_i y_i^+ + \theta_{i+q} y_i^- \right),$$

$$m_\theta^{(2)}(y_1, \ldots, y_q) = \left(\theta_0 + \sum_{i=1}^{q} \theta_i |y_i|^\delta \right)^{1/\delta},$$

where x^+ and x^- denote, respectively, the positive part and the negative part of a real number x and $\delta \geq 1$.

2. When ε_t is not necessarily independent from $\mathcal{F}_{t-1}$, (10) can be replaced with the following more abstract condition. There exists $\eta \in (0, 1)$ such that

$$\sum_{i=1}^{q} \mathbb{E}^{1/p} \left[\left(a_{i,1}(Z_0) + a_{i,2}(Z_0) |\varepsilon_1| \right)^p | \mathcal{F}_0 \right] \leq 1 - \eta \text{ a.s.} \tag{11}$$

For $q = 1$, let us compare (11) with the condition $\chi(\mathbf{A}) < 0$, which reduces to

$$\mathbb{E}\left[\log \left(a_{1,1}(Z_0) + a_{1,2}(Z_0) |\varepsilon_1| \right) \right] < 0.$$

This latter condition is much weaker than (11). Indeed, (11) entails that

$$\mathbb{E}\left[a_{1,1}(Z_0) + a_{1,2}(Z_0) |\varepsilon_1| \right] \leq \| a_{1,1}(Z_0) + a_{1,2}(Z_0) |\varepsilon_1| \|_p < 1,$$

and from Jensen's inequality, $\log \mathbb{E}\left[a_{1,1}(Z_0) + a_{1,2}(Z_0) |\varepsilon_1| \right] \leq \chi(\mathbf{A})$. For $q \geq 2$, it is more difficult to obtain explicit conditions that guaranty that $\chi(\mathbf{A}) < 0$.

3. Using the results of Liu and Wu (2010b), a nonparametric kernel estimation of the functions f and g is possible. Proposition 3 gives precise assumptions under which it is possible to control the functional dependence measure of some CHARN models when the regressors include lag values of the response as well as exogenous covariates. We then obtain additional examples of time series models for which standard nonparametric estimators of the regression function are still consistent.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

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.

$$Y_t = \varepsilon_t h_t^{1/\delta}, \quad h_t = \pi' Z_{t-1} + \sum_{i=1}^{q} \left\{ \beta_i h_{t-i} + \alpha_{i+} (Y_{t-i}^+)^\delta + \alpha_{i-} (Y_{t-i}^-)^\delta \right\}, \tag{12}$$

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

PROPOSITION 4. *Suppose that Assumptions **G1** and **G2** hold.*

1. *There then exists a unique stationary and nonanticipative solution $(Y_t)_{t\in\mathbb{Z}}$ of (12). This solution is ergodic and satisfies $\mathbb{E}|Y_0|^\delta < \infty$.*
2. *In addition, assume that Assumption **A4** holds true. Let $H_t = \left((Y_t^+)^\delta, (Y_t^+)^\delta, h_t\right)$. There then exist $C > 0$ and $\rho \in (0,1)$, such that for all $h \ge 1$, $\Theta_{1,h}(H) \le C\left[\rho^h + \sum_{i=1}^{h} \rho^i \Theta_{1,h-i}(Z)\right]$. Moreover, if $\delta \ge 1$, we have the bound*

$$\theta_{\delta,t}(Y) \le \theta_{1,t}^{1/\delta}\left(Y^+\right) + \theta_{1,t}^{1/\delta}\left(Y^-\right), \quad t \in \mathbb{N}.$$

Note

Let us consider the example of a GARCH process. We then set $\delta = 2$, $\alpha_{j+} = \alpha_{j-} = \alpha_j$, and we assume that $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a martingale difference, adapted to the filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ with $\mathcal{F}_t = \sigma\left((\varepsilon_s, Z_s) : s \le t\right)$. Set $v_{t-1} = \mathbb{E}\left[\varepsilon_t^2 | \mathcal{F}_{t-1}\right]$. If there exists a positive real number v_+ such that $v_{t-1} \le v_+$ a.s., the contraction condition in **G2** reduces to $v_+ \sum_{j=1}^{q} \left(\alpha_j + \beta_j\right) < 1$. For GARCH processes with i.i.d. innovations ε_t, we recover a standard condition ensuring the existence of a solution with a finite second moment.

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

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

a random intensity λ_t depending on past values and a covariate process. More precisely, we assume that

$$Y_t = N^{(t)}_{\lambda_t}, \quad \lambda_t = \beta_0 + \sum_{j=1}^{q} \beta_j \lambda_{t-j} + \sum_{j=1}^{q} \alpha_j Y_{t-j} + \pi' Z_{t-1}, \tag{13}$$

where $\left(N^{(t)}\right)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. Poisson processes with intensity 1, $\beta_0, \ldots, \beta_q$, $\alpha_1, \ldots, \alpha_q$ are 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, $\mathbb{E}|Z_1| < \infty$.

PROPOSITION 5.

1. *Suppose that Assumptions **PA1** and **PA2** hold. There then exists a unique nonanticipative, stationary, and ergodic process $(Y_t)_{t\in\mathbb{Z}}$ solution of (13).*
2. *In addition, if Assumption **A4** is also satisfied with $\varepsilon_t = N^{(t)}$, there then exist $C > 0$ and $\rho \in (0,1)$, such that for all $h \geq 1$, $\Theta_{1,h}((Y_t, \lambda_t)_t) \leq C\left[\rho^h + \sum_{i=1}^{h} \rho^i \Theta_{1,h-i}(Z)\right]$.*

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. Second, 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.

4.4. Dynamic Binary Choice Model

We consider the dynamic

$$Y_t = \mathbb{1}_{g(Y_{t-1},\ldots,Y_{t-q},\zeta_t)>0}, \tag{14}$$

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 the dynamics of recessions. See de Jong and Woutersen (2011) who studied the case g linear and Kauppi and Saikkonen (2008) for a study of U.S. recessions.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

PROPOSITION 6.

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

$$\mathbb{P}\left(\min_{y\in\{0,1\}^q,\,1\le t\le q} g(y,\zeta_t) > 0\right) + \mathbb{P}\left(\max_{y\in\{0,1\}^q,\,1\le t\le q} g(y,\zeta_t) \le 0\right) > 0. \tag{15}$$

There then exists a unique stationary and ergodic solution $(Y_t)_{t\in\mathbb{Z}}$ *for the recursions (14).*

2. *Assume that for some real numbers* $a_1,\dots,a_q$ *and* $\pi\in\mathbb{R}^e$, $g(y,\zeta_t)=\sum_{i=1}^q a_i y_i + \pi' Z_{t-1} + \varepsilon_t$, *with* $\zeta_t = (Z_{t-1},\varepsilon_t)$ *satisfying* ***A4*** *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*

$$\mathbb{P}\left(\phi_- + \min_{1\le t\le q} \upsilon_t > 0 | \mathcal{F}_{-K}\right) + \mathbb{P}\left(\phi_+ + \max_{1\le t\le q} \upsilon_t \le 0 | \mathcal{F}_{-K}\right) \ge \delta \text{ a.s.}, \tag{16}$$

where

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

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

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. Consider the case of g linear as in the second point of Proposition 6. In this case, de Jong and Woutersen (2011) derived existence of a unique stationary and ergodic solution when condition (16) holds true. As shown in de Jong and Woutersen (2011), condition (16) holds, in particular, when the process $(\upsilon_t)_{t\in\mathbb{Z}}$ is m-dependent or for some infinite moving averages. Condition (15) is much weaker, since it holds as soon as

$$\mathbb{P}\left(\phi_+ + \max_{1\le t\le q} \upsilon_t \le 0\right) + \mathbb{P}\left(\phi_- + \min_{1\le t\le q} \upsilon_t > 0\right) > 0. \tag{17}$$

Condition (17) holds true as soon as the random vector $(\upsilon_1,\dots,\upsilon_q)$ has full support. Another sufficient condition for (17) is the following. If $(\upsilon_t)_{t\in\mathbb{Z}}$ is adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, we assume that for any $x\in\mathbb{R}$ and $t\in\mathbb{Z}$, $\mathbb{P}(\upsilon_t \le x|\mathcal{F}_{t-1}) > 0$ a.s. or for any $t\in\mathbb{Z}$ and $x\in\mathbb{R}$, $\mathbb{P}(\upsilon_t > x|\mathcal{F}_{t-1}) > 0$ a.s. Recall that $\upsilon_t = \pi' Z_{t-1} + \varepsilon_t$. The latter condition is valid, in particular, when ε_t has full support and is independent from $\mathcal{F}_{t-1} = \sigma((\varepsilon_s, Z_s) : s \le t-1)$.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

2. As the proof of Proposition 6 will show, the condition (16) implies Assumption **A2**. Condition (15) is only used for applying Theorem 1. However, (15) does not entail mixing properties. In contrast, condition (16) does. See de Jong and Woutersen (2011, Thm. 2). Our results (see the second point of Proposition 6) give a complement when the covariate process is not necessarily strongly mixing and has a Bernoulli shift representation.
3. When $\zeta_t = (Z_{t-1}, \varepsilon_t) \in \mathbb{R}^{d+1}$ in (14), 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 $\mathcal{F}_{t-1} = \sigma\left((\varepsilon_{t-j}, Z_{t-j}) : j \geq 1\right)$, one can show, using the same arguments as in the previous point, that condition (15) is satisfied as soon as the distribution of ε_t has 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 Assumption **A2** when the covariate process $(Z_t)_{t\in\mathbb{Z}}$ is not bounded. However, when the c.d.f. 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 pseudolikelihood estimators of the parameters.

4.5. Categorical Time Series with Covariates

We consider a finite set $E = \{1, 2, \ldots, N\}$, an integer $q \geq 1$, 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 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 this kind of model but with a more general observation-driven form, such as in GARCH models, and that will not fall into 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 $u \in [0, 1]$, $z \in E_1'$, $y \in E^q$,

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

and $u \in [0,1]$, we set

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

and

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

We introduce the following assumptions.

C1 The probability kernels K_z are lower bounded by a positive constant, i.e., for any $z \in E'$, $\eta(z) := \min_{(i,y)\in E^{q+1}} K_z(i|y) > 0$.

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

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

$$\sum_{i=1}^{N} \left| K_z(i|y_1,\dots,y_q) - K_{\bar{z}}(i|y_1,\dots,y_q) \right| \leq C|z-\bar{z}|.$$

PROPOSITION 7. *Suppose that Assumptions **C1** and **C2** hold.*

1. *There exists a unique stationary process satisfying the recursions*

$$Y_t = K_{Z_{t-1}}^- \left(\varepsilon_t | Y_{t-1},\dots,Y_{t-q}\right), \quad t \in \mathbb{Z}. \tag{18}$$

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

2. *In addition, assume that Assumptions **A4** and **C3** hold true and that $\eta_- = \inf_{z\in E_1'} \eta(z) > 0$ in **C1**. There then exist $C > 0$ 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

A proof of the first point of Proposition 7 is based on Theorem 1 and provides a general result for existence of stationary categorical time series with covariates. In particular, probit, logistic, and multinomial autoregressions can be considered without restriction for the covariate process $(Z_t)_{t\in\mathbb{Z}}$. However, the derivation of the dependence properties in the second point imposes a more restrictive assumption on the transition kernel K, because it is necessary to check Assumption **A2**. Recently, Fokianos and Truquet (2019) studied categorical time series under the strict exogeneity assumption for the covariate process. For the recursions (18), strict exogeneity holds true as soon as the two processes $(Z_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ are independent. Assumption **C2** is weaker than strict exogeneity in general.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

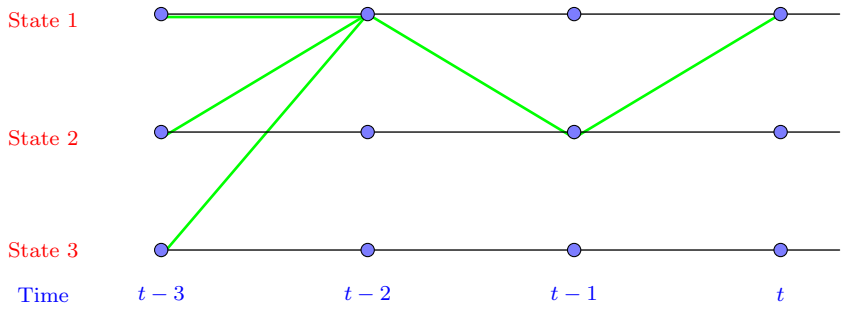


FIGURE 1. Illustration of the convergence for $N = 3$ modalities and $q = 1$ lag.

4.6. Categorical Time Series and Coalescence of the Paths

In this section, we give another interpretation of the convergence of the backward iterations for categorical time series. This interpretation has a link with some perfect simulation schemes that are widely known for Markov chains. See Propp and Wilson (1996). Since the state space is discrete, the iterations should be automatically constant after some steps. Figure 1 illustrates the convergence when $q = 1$ and $N = 3$. In this case, $f_t(j) = K^-_{Z_{t-1}}(\varepsilon_t|j)$ for $j = 1, 2, 3$. Setting

$$T = \inf\left\{k \geq 1 : \varepsilon_{t-k} \leq \min_{i,j} K_{Z_{t-k-1}}(i|j)\right\},$$

we know that from the ergodicity assumption **C2** and the positivity assumption **C1**, T is finite almost surely. In this case, $f_{t-T}(j) = 1$ a.s., $f^t_{-\infty} := \lim_{n\to\infty} f^t_{t-n}(j) = f^t_{t-T}(j)$ a.s., and the limit does not depend on the state j. All the paths corresponding to $f^t_{t-n}(j)$ for $n \geq T$ coalesce through State 1.

When $q \geq 2$, our assumptions guarantee that it is possible to get q times successively the value 1 for the time series whatever the previous values. A coalescence property for the paths will then also occur in this case.

This interpretation is also relevant for getting an interpretation of condition (15) for binary choice models. When (15) is satisfied, it is possible to get, starting at any time t, either q times the value 1 or q times the value 0 whatever the previous values of the binary time series. Running the backward iterations, we have coalescence of the paths at the first (random) time t–T such that such an event occurs.

5. CONCLUSION

A general theoretical analysis of nonlinear autoregressive time series models with exogenous covariates is absent from the present time series literature, and only a few references consider such a problem, mainly for specific examples. The aim of this paper was to provide some results for a reasonable class of nonlinear time series models for which the required assumptions can be checked.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

In particular, we provide two results, Theorems 2 and 4, which justify existence and uniqueness of stationary and ergodic solutions possessing some moments. The crucial assumption to check, **A2** or **B3**, involves a uniform conditional contraction condition. Assumption **B3** is the main assumption to check for autoregressive models with several lags. For some nonlinear models already considered in the literature, such as GARCH or autoregressive Poisson processes, this contraction condition is easily checked, because the exogenous covariates have an additive contribution in the expression of the latent process and play the role of a random intercept, which does not modify the usual stability conditions. However, our results can also be applied to autoregressive processes for which lag parameters depend on the covariates (see Section 2.4, the notes after Theorem 4, Proposition 3, and Proposition 6 for some examples). In this case, a uniform control of the random lag parameters is necessary to check our assumptions, which shows the limit of our approach.

It may be possible to weaken our uniform contraction condition, as shown in Theorem 3, at least under a strict exogeneity assumption. However, getting additional general results to ensure existence of some unconditional moments, to control dependence coefficients, and to consider higher-order autoregressive processes would require substantial effort. The proposed framework is also useful for deriving weak dependence properties of the solution, leading to the possibility to apply many existing statistical inference procedures, the central limit theorem of Section 5 providing an illustration. To this end, the functional dependence measure discussed in Section 3 is of primary importance. Note that a general result for getting weak dependence properties of autoregressive processes with exogenous covariates is also new, and it is another contribution of this paper. Finally, we also derived results for categorical time series in Sections 4.4 and 4.5. Apart from the weak dependence properties that can be derived from our general results, we also obtained stationarity conditions with weaker assumptions, applying Theorem 1. Note that whatever the results used in the paper (Theorem 1, 2, or 3), the convergence of the backward iterations of random maps appears to be a central point of view for considering many different types of autoregressive systems with exogenous regressors.

6. PROOFS OF THE RESULTS

6.1. Proof of Theorem 2

We use the convention $f_t^{t-1}(x) = x$ for $(x,t) \in E \times \mathbb{Z}$. From Assumption **A2**, if $(t,s,s') \in \mathbb{Z}^3$ such that $s' \leq s \leq t$ and $x, y \in E$,

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

and then

$$\|d\left(f_s^t(x), f_{s'}^t(y)\right)\|_p \leq L \|d\left(f_s^{t-1}(x), f_{s'}^{t-1}(y)\right)\|_p. \tag{19}$$

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

Applying (19) with $y = x_0$, $s = s' = t$, we get $\sup_{t\in\mathbb{Z}} \|d(f_t(x), f_t(x_0))\|_p < \infty$. Next, using **A1** and the triangular inequality, we get

$$\sup_{t\in\mathbb{Z}} \|d(f_t(x), y)\|_p < \infty \text{ for every } (x, y) \in E^2. \tag{20}$$

With the same kind of arguments, we get for $s \le t - m$,

$$\|d\left(f_s^t(x), f_{s'}^t(y)\right)\|_p \le \kappa \|d\left(f_s^{t-m}(x), f_{s'}^{t-m}(y)\right)\|_p. \tag{21}$$

1. We denote by $[z]$ the integer part of a real number z. To apply recursively the previous bounds, we note that for any integer i, $i+1 = r_1 m + r_2$ with $r_1 = [(i+1)/m]$ and $r_2 = i + 1 - r_1 m$. We then get from (19) and (21),

$$\begin{aligned}\sum_{i\ge 0} \|d\left(f_{t-i}^t(x), f_{t-i-1}^t(x)\right)\|_p &\le \sum_{i\ge 0} \kappa^{[(i+1)/m]} L^{i+1-m[(i+1)/m]} \|d(x, f_{t-i-1}(x))\|_p \\ &\le \frac{(L+1)^m \kappa^{(1-m)/m}}{1-\kappa^{1/m}} \sup_{j\in\mathbb{Z}} \|d\left(x, f_j(x)\right)\|_p.\end{aligned}$$

   This latter bound entails that the series $\sum_{i\ge 0} d\left(f_{t-i}^t(x), f_{t-i-1}^t(x)\right)$ 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. Moreover, from the previous bound, (20) and the triangular inequality, we deduce that

$$\sum_{s,t\in\mathbb{Z}, s\le t} \|d\left(y, f_s^t(x)\right)\|_p < \infty \text{ for every } (x, y) \in E^2. \tag{22}$$

   Next, we note that the convergence also holds in $\mathbb{L}^p$, since from Fatou's lemma,

$$\begin{aligned}\|d\big(X_t(x), f_{t-s}^t(x)\big)\|_p &\le \liminf_{j\to\infty} \|d\big(f_{t-j}^t(x), f_{t-s}^t(x)\big)\|_p \\ &\le \sum_{i\ge s} \|d\big(f_{t-i}^t(x), f_{t-i-1}^t(x)\big)\|_p \\ &\le \kappa^{s/m} \frac{(L+1)^m \kappa^{(1-m)/m}}{1-\kappa^{1/m}} \sup_{j\in\mathbb{Z}} \|d\big(f_j(x), x\big)\|_p \to 0 \text{ as } s \to \infty.\end{aligned}$$

   Finally, using (22) and the triangular inequality, we get the last assertion $\sup_{t\in\mathbb{Z}} \|d(X_t(x), x_0)\|_p < \infty$.
2. If $x \ne y$, we have from the almost sure convergence and Fatou's lemma,

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

   This shows the second point.
3. For the third point, we observe that for any $j \ge 1$, there exists a measurable function $H_j^{(x)} : E'^{j+1} \to E$ such that $f_{t-j}^t(x) = H_j^{(x)}(\zeta_t, \dots, \zeta_{t-j})$. Since $\lim_{j\to\infty} H_j^{(x)}$ exists $\mathbb{P}_{(\zeta_{t-j})_{j\ge 0}}$ a.s., it is then possible to define a measurable function $H : (E')^{\mathbb{N}} \to E$ such that $X_t = H((\zeta_{t-j})_{j\ge 0})$ a.s. The process $(X_t)_{t\in\mathbb{Z}}$ has a Bernoulli

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

shift structure with dependent entries and is then stationary and ergodic provided that 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 $j \geq 1$:

$$\|d(X_t, Y_t)\|_p = \|d\left(f^t_{t-mj+1}(X_{t-mj}), f^t_{t-mj+1}(Y_{t-mj})\right)\|_p$$
$$\leq \kappa^j \left[\sup_{j\in\mathbb{Z}} \|d(x_0, X_j)\|_p + \sup_{j\in\mathbb{Z}} \|d(x_0, Y_j)\|_p\right]. \qquad \square$$

6.2. Proof of Theorem 3

1. From Assumption **A0**, $f^t_{t-s}(x)$ is, conditionally on Z, an iteration of $s+1$ independent random maps. Using Assumption **A2'**, we get

$$\sum_{s\geq 0} \mathbb{E}^{1/p}\left[d^p\left(f^t_{t-s}(x), f^t_{t-s-1}(x)\right)|Z\right] \leq \sum_{s\geq 0}\prod_{i=1}^{s+1} \kappa(Z_{t-i}) b_{t-s-2}(x),$$

with

$$b^p_t(x) = \int d^p\left(x, F(x, Z_t, u)\right) d\mathbb{P}_{\varepsilon_0}(u).$$

From Assumption **A2'** and the triangular inequality, we have $\mathbb{E}\log^+ b_t(x) < \infty$ for any $x \in E$. We are going to show that

$$\sum_{s\geq 0} \mathbb{E}^{1/p}\left[d^p\left(f^t_{t-s}(x), f^t_{t-s-1}(x)\right)|Z\right] < \infty \text{ a.s.} \qquad \textbf{(23)}$$

This follows from the assumptions on the logarithmic moments. Indeed, $((\kappa(Z_t), b_t(x)))_{t\in\mathbb{Z}}$ is a stationary process, and it is widely known that the stochastic recursions

$$Y_t = \kappa(Z_{t-1})Y_{t-1} + b_{t-1}(x)$$

have a unique stationary solution given by

$$Y_t = b_{t-1}(x) + \sum_{s\geq 0}\prod_{i=1}^{s+1} \kappa(Z_{t-i}) b_{t-s-2}(x),$$

the latter series being convergent almost surely. See for instance Brandt (1986, Thm. 1). This shows (23). Using Minkowski's inequality and Fatou's lemma for conditional expectations, see for instance Kallenberg (2006, Chap. 5), we then deduce that

$$\mathbb{E}^{1/p}\left[S(x)^p|Z\right] < \infty \text{ a.s. } S(x) := \sum_{s\geq 0} d\left(f^t_{t-s}(x), f^t_{t-s-1}(x)\right).$$

As a consequence, we have $\mathbb{P}\left(S(x) < \infty|Z\right) = 1$ a.s. and then $\mathbb{P}\left(S(x) < \infty\right) = 1$. From the Cauchy criterion, we then conclude the existence of a random variable $X_t(x)$ such that $\lim_{s\to\infty} f^t_{t-s}(x) = X_t(x)$ a.s. Note that from Assumption **A1'** and

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

(23), we have $\mathbb{E}\left[d^p\left(f^t_{t-s}(x),x_0\right)|Z\right] < \infty$ a.s. for every positive integer s. The convergence $\lim_{s\to\infty}\mathbb{E}\left[d^p\left(f^t_{t-s}(x),X_t(x)\right)|Z\right]=0$ a.s. and $\mathbb{E}\left[d^p\left(X_t(x),x_0\right)|Z\right] < \infty$ a.s. follow as in the proof of Theorem 2, using Fatou's lemma for conditional expectation.

2. For a positive integer s, we have from **A2**,

$$\mathbb{E}^{1/p}\left[d^p\left(f^t_{t-s}(x),f^t_{t-s}(y)\right)|Z\right] \le \prod_{i=1}^{s}\kappa(Z_{t-i})d(x,y) \to 0 \text{ a.s.}$$

Letting $s\to\infty$, we have $\mathbb{E}[d^p(X_t(x),X_t(y))|Z]=0$ a.s. and then $\mathbb{P}(X_t(x)\neq X_t(y)|Z)=0$ a.s. Taking the expectation, we conclude that $\mathbb{P}(X_t(x)\neq X_t(y))=0$.

3. From the almost sure convergence of the sequence $(f^t_{t-s}(x))_{s\ge 0}$, stationarity and ergodicity of the process $((X_t,Z_t))_{t\in\mathbb{Z}}$ follow exactly as in the proof of the third point 3 Theorem 2.

4. Let $(Y_t)_{t\in\mathbb{Z}}$ be a stochastic process satisfying the proposed conditions. If the process is nonanticipative, we have from **A2'**,

$$\begin{aligned}\mathbb{E}[d^p(X_t,Y_t)|Z] &= \mathbb{E}[d^p(f^t_{t-s}(X_{t-s-1}),f^t_{t-s}(Y_{t-s-1}))|Z]\\ &\le \prod_{i=0}^{s}\kappa^p(Z_{t-i-1})\cdot\mathbb{E}[d^p(X_{t-s-1},Y_{t-s-1})|Z].\end{aligned}$$

Note that, from **A2'**, $\prod_{i=0}^{s}\kappa^p(Z_{t-i-1})=o_{\mathbb{P}}(1)$. Moreover,

$$\begin{aligned}&\mathbb{E}^{1/p}[d^p(X_{t-s-1},Y_{t-s-1})|Z]\\ &\le \mathbb{E}^{1/p}[d^p(X_{t-s-1},x_0)|Z]+\mathbb{E}^{1/p}[d^p(x_0,Y_{t-s-1})|Z].\end{aligned}$$

Note that

$$\mathbb{E}^{1/p}[d^p(x_0,Y_{t-s-1})|Z]=\mathbb{E}^{1/p}[d^p(x_0,Y_{t-s-1})|Z_{t-s-1},Z_{t-s-2},\ldots]$$

and if the process $((Z_t,Y_t))_{t\in\mathbb{Z}}$ is stationary, then the process $(V_t)_{t\in\mathbb{Z}}$ defined by

$$V_t=\mathbb{E}^{1/p}\left[d^p\left(x_0,Y_t\right)|Z_t,Z_{t-1},\ldots\right]$$

is also stationary and takes finite values from our assumptions. Then, $V_{t-s-1}=O_{\mathbb{P}}(1)$, where s is the index of the sequence. The same property holds if Y_t is replaced with X_t. As a consequence

$$\mathbb{E}\left[d^p\left(X_{t-s-1},Y_{t-s-1}\right)|Z\right]=O_{\mathbb{P}}(1).$$

We then conclude that $\mathbb{E}[d^p(X_t,Y_t)|Z]=0$ a.s. Then, $\mathbb{P}[X_t\neq Y_t|Z]=0$ a.s. and by integration, we get the conclusion.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

6.3. Proof of Proposition 1

1. We use the decomposition

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

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

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

If $i = t-1$, we have $\| d\left(f^t_{t-i} \circ \overline{f}^{t-i-1}_{-\infty}, f^t_{t-i-1} \circ \overline{f}^{t-i-2}_{-\infty} \right) \|_p \le 2 \| d\left(0, f^0_{-\infty}\right) \|_p \kappa^{t/m-1}$. Using the triangular inequality, we get for $t \ge 2$,

$$\begin{aligned} \theta_{p,t} &\le \kappa^{-1} L^m \| S(f^0_{-\infty}) \|_s \sum_{i=0}^{t-2} \kappa^{(i+1)/m} \| H_{t-i-2} \|_r \\ &\quad + \| S(f^0_{-\infty}) \|_s \| H_{t-1} \|_r + 2\kappa^{t/m-1} \| d\left(0, f^0_{-\infty}\right) \|_p. \end{aligned}$$

The bound (7) is obtained by summation and entails the simpler bound (8).

2. From **A4'**, we have **A4** with $H_{t-1} = d\left(Z_{t-1}, \overline{Z}_{t-1}\right)$ with

$$\overline{Z}_t = H'\left(\eta_t, \ldots, \eta_1, \eta'_0, \eta_{-1}, \ldots\right).$$

We then deduce the result from the previous point, noticing that $\eta_{r,j} = \Theta_{r,j-1}(Z)$.

6.4. Proof of Theorem 4

Define the following random map

$$f_t(u_1, \ldots, u_q) = \left(F\left(u_1, \ldots, u_q, \zeta_t\right)', u'_1, \ldots, u'_{q-1} \right)'.$$

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

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

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

$$\left(U_t(x), \ldots, U_{t-q+1}(x)\right) = f^t_{q+1}(x).$$

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

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

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

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 $v = (v_1', \ldots, v_q')' \in \mathbb{R}^{kq} \setminus \{0\}$ 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

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

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

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

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

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

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

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

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 $d = \overline{d}$, $m = \inf\{j \geq 1 : \left|\mathbb{1}'B^{j}\right|_{\infty} < 1\}$, and $\kappa^{p} = \left|\mathbb{1}'B^{m}\right|_{\infty}$. Moreover, **A1** is a direct consequence of **B2**. The result then follows from Theorem 2. □

6.5. Proof of Proposition 2

Defining

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

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

we set for $t \geq 0$, $\overline{Z}_t = H\left(\eta_t, \ldots, \eta_1, \eta_0', \eta_{-1}, \ldots\right)$. Using Assumptions **A5** and **B5**, we have, for $t \geq 1$,

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

We then check Assumption **A4**, setting $H_{t-1} = d\left(Z_{t-1},\overline{Z}_{t-1}\right)$. One can then apply Proposition 1 using, for instance, the ℓ_1 distance on E^q defined by $\overline{d}(u,v) = \sum_{i=1}^q d(u_i,v_i)$, which is equivalent to the distance $\overline{d}$ used in the proof of Theorem 4. Since $\mathbb{E}^{1/r}\left[d^r\left(Z_{t-1},\overline{Z}_{t-1}\right)\right] = \theta_{r,t-1}(Z)$, the proof of the proposition is now complete.

6.6. Proof of Theorem 5

The proof is a consequence of Thm. 3 in Wu (2005). It is simply necessary to prove that

$$\Gamma(Y) = \sum_{t=0}^{\infty} \|Y_t - \overline{Y}_t\|_2 < \infty.$$

To this end, for $t \geq 0$, let $\overline{X}_t$ be the random variables obtained by replacing ξ_0 by an independent copy ξ_0' in its Bernoulli shift representation. We then have $\overline{Y}_t = f\left(\overline{X}_t, \ldots, \overline{X}_{t-k}\right)$ for $t \geq k$. From the assumption on the function f, stationarity and Hölder's inequality, we have, setting $\ell_1 = 2p/(p-2)$,

$$\theta_{2,t}(Y) \leq C\left(1 + 2k\||X_0|^{o\ell}\|_{\ell_1}\right) \cdot \sum_{i=0}^{k} \theta_{p,t-i}(X).$$

The required condition easily follows by summation.

6.7. 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:

   $$g_t(x) = \left(f_1(x_1,\ldots,x_q,Z_{t-1}) + \varepsilon_t f_2(x_1,\ldots,x_q,Z_{t-1}), x_1, \ldots, x_{q-1}\right)'.$$

   We then have

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

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

   $$\left|g_1^t(x) - g_1^t(y)\right|_{vec} \preceq A_t \cdots A_1 \cdot |x-y|_{vec}.$$

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

We then deduce that $c(g_1^t) \leq \|A_t \cdots A_1\|$. The result is then a consequence of Theorem 1, using the condition $\chi(\mathbf{A}) < 0$.

2. We check the assumptions of Theorem 4. First note that from Assumption **CH1**, the process $\zeta_t = (Z_{t-1}, \varepsilon_t)$ is ergodic. This entails **B1**. Next, we set

$$F(x_1, \ldots, x_q, \zeta_t) = f_1(x_1, \ldots, x_q, Z_{t-1}) + \varepsilon_t f_2(x_1, \ldots, x_q, Z_{t-1}).$$

Our assumptions guarantee that $F(\bar{x}, \zeta_1) \in \mathbb{L}^p$ 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**. To this end, for $i = 1, \ldots, q$, we set $\delta_i = \sup_z \|a_{i,1}(z) + a_{i,2}(z)|\varepsilon_1|\|_p$. Using Minkowski's inequality for conditional expectations (see for instance Doob, 2012, Chap. 11, Sect. 3), we have

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

Next, using convexity, we get

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

B3 is then a consequence of **CH3**.

3. We apply Proposition 2. From the previous points, it is only required to check **B4**, which is a consequence of Assumption **CH3**.

6.8. Proof of Proposition 4

For the first part, we apply Theorem 4. 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

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

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

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** and **G2**.

For the second part, it is easily seen that **B4** is satisfied for a constant function S, $p = 1$ and $s = \infty$. This gives the bound for $\Theta_{1,h}(H)$. If $\delta \geq 1$, the last bound for $\theta_{\delta,t}(Y)$ can be obtained from the inequalities

$$|x - y|^{\delta} \leq \left|x^{\delta} - y^{\delta}\right|, \quad x, y \geq 0.$$

6.9. Proof of Proposition 5

1. To show the first point, we check the assumptions of Theorem 4. 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 (13) if and only if $X_t = (Y_t, \lambda_t)'$ is the solution of $X_t = F\left(X_{t-1}, \ldots, X_{t-q}, \zeta_t\right)$ with

$$F\left(x_1, \ldots, x_q, \zeta_t\right) = \left(N^{(t)}_{f(x_1,\ldots,x_q,Z_{t-1})}, f(x_1, \ldots, x_q, Z_{t-1})\right)'$$

and $f(x_1, \ldots, x_q, Z_{t-1}) = \beta_0 + \sum_{j=1}^{q} \beta_j s_j + \sum_{j=1}^{q} \alpha_j y_j + \pi' Z_{t-1}$, $x_i = (y_i, s_i)$, $1 \leq i \leq q$. For $x \in (\mathbb{N} \times \mathbb{R}_+)^q$,

$$\mathbb{E}\left[|F(x, \zeta_1)|\right] = 2\left(\beta_0 + \sum_{j=1}^{q} \beta_j s_j + \sum_{j=1}^{q} \alpha_j y_j + \pi' \mathbb{E}(Z_1)\right) < \infty,$$

since $\mathbb{E}(|Z_1|) < \infty$.

We then have, for $(x, x') \in ((\mathbb{N} \times \mathbb{R}_+)^q)^2$ with $x = (x_1, \ldots, x_q), x' = (x'_1, \ldots, x'_q)$ $\forall j = 1, \ldots, q, x_j = (y_j, s_j), x_j = (y'_j, s'_j)$,

$$\mathbb{E}\left[\left|F(x, \zeta_t) - F(x', \zeta_t)\right|_{vec} | \mathcal{F}_{t-1}\right] \preceq \sum_{j=1}^{q} \begin{pmatrix} \alpha_j & \beta_j \\ \alpha_j & \beta_j \end{pmatrix} \left|x_j - x'_j\right|_{vec}.$$

In the previous bounds, we have used the identity $\mathbb{E}\left[|N^{(t)}_{h_{t-1}} - N^{(t)}_{g_{t-1}}| \middle| \mathcal{F}_{t-1}\right] = |h_{t-1} - g_{t-1}|$, which is valid for two nonnegative random variables h_{t-1} and g_{t-1} measurable with respect to $\mathcal{F}_{t-1}$. The previous equality follows from the

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

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 4 are satisfied. Hence, according to Theorem 4, there exists a unique stationary and nonanticipative process $(X_t)_{t\in\mathbb{Z}}$ solution of (13) such that $\mathbb{E}[|X_t|] < \infty$. 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 **B4** for $p = r = 1$ and $s = \infty$. This is straightforward, since we have the equality

$$\mathbb{E}\left[|F(x_1,\dots,x_q,z,N^{(t)}) - F(x_1,\dots,x_q,\bar{z},N^{(t)})|\right] = 2|\pi'(z-\bar{z})|.$$

The proof of the second point is now complete.

6.10. 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

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

We set

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

Setting $\left(y_t(x),\dots,y_{t-q+1}(x)\right)' = f_1^t(x)$ for $t \ge q$, we have

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

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

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

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

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

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

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

Setting $t = 1$, this shows, in particular, that

$$c\left(f_1^q\right) \le \max_{1\le i\le q} \delta_i \le \mathbb{1}_{\max_{y,i} g(y,\zeta_i)>0} - \mathbb{1}_{\min_{y,i} g(y,\zeta_i)>0}. \tag{24}$$

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

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 (24) and the inequality $h_s \le 1$ for all $s \in \mathbb{Z}$ to get

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

Moreover, using (16), we have

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

provided that $Jq \ge q + K$. This guarantees **A2**, with $\kappa = 1-\delta$ and $m = Jq$. Finally, let us check **A3**. Setting $\overline{Z}_t = H\left(\eta_t, \dots, \eta_1, \eta_0', \eta_{-1}, \dots\right)$. We have

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

where L_ε denotes the Lipschitz constant of F_ε.

6.11. Proof of Proposition 7

1. For the first point, we will apply Theorem 1 for the discrete metric $d(x,y) = \mathbb{1}_{x\ne y}$. Setting for $x \in E^q$ and $t \in \mathbb{Z}$, $f^t_{t-q+1}(x) = f_t \circ \cdots \circ f_{t-q+1}(x)$, we have, for $x, y \in E^q$,

$$\begin{aligned}\left\{f^t_{t-q+1}(x) = f^t_{t-q+1}(y)\right\} &\supset \left\{f^t_{t-q+1}(x) = f^t_{t-q+1}(y) = (1,\dots,1)\right\}\\ &= \left\{\varepsilon_j \in [0, \eta(Z_{j-1})] : j = t-q+1, \dots t\right\}.\end{aligned}$$

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

We then obtain

$$d\left(f^t_{t-q+1}(x),f^t_{t-q+1}(y)\right) \le \left(1 - \prod_{j=t-q+1}^{t} \mathbb{1}_{\varepsilon_j \in [0,\eta(Z_{j-1})]}\right) d(x,y).$$

Let us show that

$$p := \mathbb{P}\left(\varepsilon_j \in [0,\eta(Z_{j-1})] : j = t-q+1,\dots t\right) > 0. \tag{25}$$

Note that by stationarity, p does not depend on t. Assume that $p = 0$. From Assumption **C2** and the properties of the conditional expectations, we have

$$p = \mathbb{E}\left[\eta(Z_{t-1}) \prod_{j=t-q+1}^{t-1} \mathbb{1}_{\{\varepsilon_j \le \eta(Z_{j-1})\}}\right] = 0.$$

Since η is positive, we get $\mathbb{P}\left(\varepsilon_j \in [0,\eta(Z_{j-1})] : j = t-q+1,\dots t-1\right) = 0$. By finite induction, we deduce that $\mathbb{P}\left(\varepsilon_{t-q+1} \le \eta(Z_{t-q})\right) = 0$. Since this latter probability equals to $\mathbb{E}\left(\eta(Z_{t-q})\right) > 0$, we obtain a contradiction. The property (25) is then valid and Theorem 1 applies, which leads to the conclusion.

2. For the second point, we will use Theorem 2 and Proposition 1. Since d is a bounded metric, **A1** is automatically satisfied. Next, observe that

$$\begin{aligned}
\mathbb{P}\left(f^t_{t-q+1}(x) = f^t_{t-q+1}(y)|\mathcal{F}_{t-q}\right) &\ge \mathbb{P}\left(f^t_{t-q+1}(x) = f^t_{t-q+1}(y) = (1,\dots,1)|\mathcal{F}_{t-q}\right) \\
&\ge \mathbb{P}\left(\varepsilon_t,\dots,\varepsilon_{t-q+1} \in [0,\eta_-]\mathcal{F}_{t-q}\right) \\
&= \mathbb{P}\left(\varepsilon_t,\dots,\varepsilon_{t-q+1} \in [0,\eta_-]\right) \\
&\ge \eta_-^q.
\end{aligned}$$

This yields to the bound $\mathbb{E}\left[d\left(f^t_{t-q+1}(x),f^t_{t-q+1}(y)|\mathcal{F}_{t-q}\right)\right] \le 1-\eta_-^q$, which shows the second part of **A2**. The first part is automatic.

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

$$\begin{aligned}
&\mathbb{E}\left[|K_z^-(\varepsilon_1|y_1,\dots,y_q) - K_{z'}^-(\varepsilon_1|y_1,\dots,y_q)|\right] \\
&\quad\le \int_0^1 \left|K_z^-(u|y_1,\dots,y_q) - K_{z'}^-(u|y_1,\dots,y_q)\right| du \\
&\quad\le \sum_{j=1}^{N} \left|\sum_{i=1}^{j} K_z(i|y_1,\dots,y_q) - \sum_{i=1}^{j} K_{z'}(i|y_1,\dots,y_q)\right| \\
&\quad\le NC\left|z-z'\right|.
\end{aligned}$$

Assumption **A3'** then follows with $s=\infty$ and $r=p=1$. The bound for the functional dependence coefficients is then a direct consequence of Proposition 1. ■

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

REFERENCES

Agosto, A., G. Cavaliere, D. Kristensen, & A. Rahbek (2016) Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX). *Journal of Empirical Finance* 38, 640–663.

Andrews, D.W.K. (1984) Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21(4), 930–934.

Borovkov, A.A. (1998) *Ergodicity and Stability of Stochastic Processes*. Wiley.

Bougerol, P. & N. Picard (1992) Strict stationarity of generalized autoregressive processes. *The Annals of Probability* 20(4), 1714–1730.

Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* 18(1), 211–220.

Chamberlain, G. (1982) The general equivalence of Granger and Sims causality. *Econometrica: Journal of the Econometric Society* 50(3), 569–581.

Davis, R.A. & H. Liu (2016) Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica* 26(4), 1673–1707.

de Jong, R.M. & T. Woutersen (2011) Dynamic time series binary choice. *Econometric Theory* 27(4), 673–702.

Diaconis, P. & D. Freedman (1999) Iterated random functions. *SIAM Review* 41(1), 45–76.

Doob, J.L. (2012) *Measure Theory*, vol. 143. Springer Science & Business Media.

Doukhan, P. (1994) *Mixing: Properties and Examples*. Lecture Notes in Statistics, vol. 85. Springer-Verlag.

Elton, J.H. (1990) A multiplicative ergodic theorem for Lipschitz maps. *Stochastic Processes and Their Applications* 34(1), 39–47.

Engle, R.F., D.F. Hendry, & J.-F. Richard (1983) Exogeneity. *Econometrica: Journal of the Econometric Society* 51(2), 277–304.

Fokianos, K. & B. Kedem (2003) Regression theory for categorical time series. *Statistical Science* 18(3), 357–376.

Fokianos, K. & L. Truquet (2019) On categorical time series models with covariates. *Stochastic Processes and Their Applications* 129(9), 3446–3462.

Francq, C. & L.Q. Thieu (2019) QML inference for volatility models with covariates. *Econometric Theory* 35(1), 37–72.

Härdle, W. & A. Tsybakov (1997) Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics* 81(1), 223–242.

Härdle, W., A. Tsybakov, & L. Yang (1998) Nonparametric vector autoregression. *Journal of Statistical Planning and Inference* 68(2), 221–245.

Iosifescu, M. (2009) *Iterated Function Systems: A Critical Survey*. Mathematical Reports (Bucureşti). 3.

Kallenberg, O. (2006) *Foundations of Modern Probability*. Springer Science & Business Media.

Kauppi, H. & P. Saikkonen (2008) Predicting US recessions with dynamic binary response models. *The Review of Economics and Statistics* 90(4), 777–791.

Letac, G. (1986) A contraction principle for certain Markov chains and its applications. *Contemporary Mathematics* 50, 263–273.

Liu, W. & W.B. Wu (2010a) Asymptotics of spectral density estimates. *Econometric Theory* 26(4), 1218–1245.

Liu, W. & W.B. Wu (2010b) Simultaneous nonparametric inference of time series. *Annals of Statistics* 38(4), 2388–2421.

Lütkepohl, H. (2005) *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media.

Pedersen, R.S. & A. Rahbek (2018) Testing GARCH-X type models. *Econometric Theory* 35(5), 1012–1047.

Propp, J.G. & D.B. Wilson (1996) Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures & Algorithms* 9(1–2), 223–252.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press

Russell, J.R. & R.F. Engle (2005) 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.

Sims, C.A. (1972) Money, income, and causality. *The American Economic Review* 62(4), 540–552.

Stenflo, Ö. (2001) Markov chains in random environments and random iterated function systems. *Transactions of the American Mathematical Society* 353(9), 3547–3562.

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

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

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

Wu, W.B., Y. Huang, & Y. Huang (2010) Kernel estimation for time series: An asymptotic theory. *Stochastic Processes and Their Applications* 120(12), 2412–2431.

Wu, W.B. & X. Shao (2004) Limit theorems for iterated random functions. *Journal of Applied Probability* 41(2), 425–436.

Wu, W.B. & Y.N. Wu (2016) Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics* 10(1), 352–379.

Xiao, H. & W.B. Wu (2012) Covariance matrix estimation for stationary time series. *The Annals of Statistics* 40(1), 466–493.

https://doi.org/10.1017/S0266466620000559 Published online by Cambridge University Press