

Geometric β -mixing of mixed causal–noncausal processes

JORDI LLORENS-TERRAZAS*, JULIEN PEIGNON†, ARTHUR THOMAS‡

October 20, 2026

Abstract

Mixed causal–noncausal autoregressive (MAR) processes are strictly stationary linear processes whose dynamics in calendar time are nonlinear, with local trends, bubbles and spikes. This nonlinearity makes their dependence properties difficult to establish. Strong mixing, or α -mixing, follows from results on linear processes, but absolute regularity, or β -mixing, was not available for mixed causal–noncausal models. We prove the geometric β -mixing of $\text{MAR}(r, s)$ processes with *i.i.d.* errors, for any number of noncausal roots, real or complex, and under a moment condition of arbitrarily small order on the errors, which allows for infinite mean and infinite variance. The proof shows that the Markov chain of their nonlinear state space representation is V -geometrically ergodic. To this end, we develop a novel Lyapunov function, which weights each future shock by the time remaining before it is realized, and verify a Foster–Lyapunov drift condition for a state that is an infinite moving average of future shocks. The drift condition implies that explosive episodes have durations with geometric tails, a nonasymptotic counterpart of known results on the survival of bubbles. The results extend to mixed causal–noncausal vector autoregressive (VAR) processes, and to MAR models with exogenous regressors or moving average components. Finally, we discuss the consequences for inference and prediction.

Keywords: Mixed causal–noncausal autoregression, Nonlinear state space representation, Geometric ergodicity, β -mixing, Prediction

*Universidad Carlos III de Madrid, jordi.llorens@uc3m.es.

†CEREMADE, CNRS UMR 7534, Université Paris-Dauphine, PSL University. Corresponding author (julien.peignon@dauphine.psl.eu).

‡Université Paris Dauphine, Université PSL, LEDA, CNRS, IRD, 75016 PARIS, FRANCE, arthur.thomas@dauphine.psl.eu.

1 Introduction

Mixed causal–noncausal autoregressive (MAR) models allow the autoregressive polynomial to have roots both inside and outside the unit circle. They have become a standard tool for modeling locally explosive patterns, such as bubbles and spikes, in economic and financial time series (Lanne and Saikkonen, 2011; Fries and Zakoian, 2019; Cavaliere et al., 2020; Davis and Song, 2020; Hecq et al., 2020). A $\text{MAR}(r, s)$ process combines a causal part with r lags, which looks back at past values, and a noncausal part with s leads, which looks ahead to future ones. Its moving average representation is therefore two-sided. The value at date t depends both on shocks that have already happened and on shocks that are still to come, so the errors are not the innovations of the process. The two parts can be told apart only when the errors are non-Gaussian (Rosenblatt, 2000), and in applications they are typically heavy-tailed. In calendar time, the process is Markov of order $r + s$ (Fries and Zakoian, 2019), but its dynamics are nonlinear. A large future shock produces a phase of explosive growth, which ends abruptly once the shock has been realized (Gouriéroux and Zakoian, 2017).

The asymptotic theory of estimators, tests, resampling methods and prediction procedures for dependent data relies on mixing conditions. For linear processes, strong mixing, or α -mixing, holds under a smoothness condition on the error density, whether the moving average representation is one-sided or two-sided (Doukhan, 1994; Rosenblatt, 2000). Absolute regularity, or β -mixing, is a stronger property, and it is the one required by blocking arguments, through the coupling lemma of Berbee (1979). These arguments underlie empirical process theory and the statistics of extremes for dependent data. For linear processes, however, β -mixing is established only in the causal case (Pham and Tran, 1985; Mokkadem, 1988). Doukhan (1994) notes that no sufficient condition was known in the noncausal case. This paper fills the gap by establishing the geometric β -mixing of MAR processes.

Building on the nonlinear state space representation of Gouriéroux and Jasiak (2026), we show that the state vector of the $\text{MAR}(r, s)$ process is a V -geometrically ergodic Markov chain (Meyn and Tweedie, 2009), and hence geometrically β -mixing. The main difficulty is to establish the Foster–Lyapunov drift condition. For a causal autoregression, the size of the state is a natural Lyapunov function. Past shocks fade at a geometric rate and each new shock is independent of the past, so a large state is pulled back toward zero on average. For a MAR process neither holds. The noncausal part of the dynamics is expansive, and the shock that moves the chain from one date to the next is itself a component of the current state. A large state may keep growing for a while, until the large shock that produced it has been realized, and its norm need not contract.

The Lyapunov functions used for nonlinear time series, which are functions of the size of the state (Liebscher, 2005; Meitz and Saikkonen, 2008; Meyn and Tweedie, 2009), therefore cannot be used here as soon as $s \geq 2$. Ours starts from the observation that

the noncausal part of the state is a linear combination of future shocks, so that the state carries information on the shocks still to come. A large state is typically produced by a single large future shock (Hult and Samorodnitsky, 2008), whose size can be inferred from the state, but whose horizon, that is, the number of periods before it is realized, may remain uncertain. Our Lyapunov function is therefore not a function of the size of the state but of the conditional law, given the state, of these future shocks. For each horizon, it takes the expected part of the future shock at that horizon above a threshold growing geometrically with the horizon, and weights it by a factor that also grows geometrically with the horizon, that is, with the time remaining before the shock is realized. After one period, every pending shock has moved one horizon closer, so its weight shrinks by a fixed factor, whatever happens to the explosive episode. The drift condition comes from the passage of time rather than from mean reversion.

Our contributions are the following. First, we show that, under mild conditions on the error density, the state vector of every $\text{MAR}(r, s)$ process with $r \geq 0$ and $s \geq 1$ is V -geometrically ergodic, and that the process is geometrically β -mixing, whatever the number, moduli and multiplicities of its noncausal roots, real or complex, and whatever the tail index of the errors, including errors with infinite variance or infinite mean. To the best of our knowledge, these are the first β -mixing results for linear processes with a genuinely two-sided moving average representation. The purely noncausal case $r = 0$ was already covered, since its time reversal is a causal autoregression and the β -mixing coefficients of a stationary process do not change under time reversal. Second, the drift condition has a direct reading in terms of explosive episodes. From any state, the time needed to return to a bounded neighborhood of the origin has geometric tails, and its expectation is at most logarithmic in the Lyapunov function. This is a nonasymptotic counterpart of the results of Gouriéroux and Zakoïan (2017), Fries (2022), Gouriéroux et al. (2025) and de Truchis et al. (2025) on the lifetime of bubbles, which describe the process conditionally on an extreme level, whereas our bound holds at every state. Third, the results extend to mixed causal–noncausal vector autoregressive (VAR) processes, to MAR models with exogenous regressors (MARX) and to MAR models with moving average components (MARMA).

These results have consequences for inference and prediction. Procedures that only require strong mixing were already within reach of the results on linear processes. The β -mixing property adds the blocking arguments, and provides the mixing part of the conditions under which the Hill estimator and estimators of cluster functionals, such as the extremal index, are analyzed for dependent data (Drees, 2000; Drees and Rootzén, 2010), a first step toward inference on the clusters of extremes of MAR processes (Fries and Zakoïan, 2019). Blocking arguments also give generalization bounds for forecasting MAR processes by empirical risk minimization, with an effective sample size of order $n/\log n$. This opens the way to a convergence proof for the mixture density networks of Dumitrescu et al. (2026).

The paper is organized as follows. [Section 2](#) introduces the model and the assumptions, and recalls the nonlinear state space representation. [Section 3](#) establishes geometric ergodicity and geometric β -mixing, and [Section 4](#) outlines the proof of the drift condition. [Section 5](#) discusses the implications for explosive episodes, inference and prediction, and [Section 6](#) extends the results to multivariate processes, exogenous regressors and moving average components. All proofs are collected in [Section A](#).

2 Model and state space representation

This section sets out the model and the assumptions, and recalls the nonlinear state space representation on which the Markov chain arguments of [Section 3](#) rest.

2.1 Model and assumptions

Let $(\varepsilon_t)_{t \in \mathbb{Z}}$ be a sequence of independent and identically distributed (*i.i.d.*) real random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We consider the MAR(r, s) equation

$$\psi(F)\phi(B)X_t = \varepsilon_t, \quad t \in \mathbb{Z}, \quad (2.1)$$

where

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_r B^r, \quad \psi(F) = 1 - \psi_1 F - \dots - \psi_s F^s.$$

Here B and F denote the backward and forward shift operators, $BX_t = X_{t-1}$ and $FX_t = X_{t+1}$. The integers $r \geq 0$ and $s \geq 0$ are the causal and noncausal orders, and $\phi(z)$, $\psi(z)$, $z \in \mathbb{C}$, denote the associated polynomials, with real coefficients.

We impose the following three assumptions. [Assumption 1](#) concerns the orders and the roots of the polynomials ϕ and ψ , while [Assumptions 2](#) and [3](#) concern the regularity and the tails of the error density.

Assumption 1. *The orders and the polynomials satisfy:*

(ι) $s \geq 1$ and $\psi_s \neq 0$; if $r \geq 1$, $\phi_r \neq 0$;

(υ) all roots of $\phi(z)$ and $\psi(z)$ lie strictly outside the unit circle, that is, $\phi(z)\psi(z) \neq 0$ for every $z \in \mathbb{C}$ with $|z| \leq 1$.

Assumption 2. *The error term ε_t admits a density f_ε with respect to the Lebesgue measure on $\mathbb{R}$, that is positive, continuous and bounded on $\mathbb{R}$.*

Assumption 3. *There exist $\alpha > 0$, a slowly varying function ℓ and constants $c_+ > 0$, $c_- > 0$ such that the density f_ε of [Assumption 2](#) satisfies*

$$f_\varepsilon(x) \sim c_\pm |x|^{-(\alpha+1)} \ell(|x|), \quad x \rightarrow \pm\infty,$$

where $a(x) \sim b(x)$ means $a(x)/b(x) \rightarrow 1$.

Condition (ι) of [Assumption 1](#) ensures that the orders are exactly r and s , and that the process has at least one noncausal root. Under condition (ι) , let $\lambda_\phi > 1$ and $\lambda_* > 1$ denote the smallest moduli of the roots of ϕ and ψ , respectively, with the convention $\lambda_\phi = +\infty$ if $r = 0$. Since $\psi(z^{-1}) = 0$ implies $|z| \leq \lambda_*^{-1}$ and $\phi(z) = 0$ implies $|z| \geq \lambda_\phi$, the function $z \mapsto 1/\{\phi(z)\psi(z^{-1})\}$ is holomorphic on the annulus $\mathbb{A} = \{z \in \mathbb{C} : \lambda_*^{-1} < |z| < \lambda_\phi\}$, which contains the unit circle, and admits there the Laurent expansion

$$\frac{1}{\phi(z)\psi(z^{-1})} = \sum_{j \in \mathbb{Z}} c_j z^j, \quad z \in \mathbb{A},$$

see [de Truchis and Thomas \(2026\)](#). By the Cauchy estimates for Laurent coefficients, for every $\chi \in (\max(\lambda_*^{-1}, \lambda_\phi^{-1}), 1)$ there exists $C_\chi < \infty$ such that $|c_j| \leq C_\chi \chi^{|j|}$ for all $j \in \mathbb{Z}$ ([de Truchis and Thomas, 2026](#)).

Since the slowly varying function ℓ of [Assumption 3](#) satisfies $\ell(x) \leq x^\delta$ for x large enough and every $\delta > 0$ ([Mikosch and Wintenberger, 2024](#)), and f_ε is bounded, [Assumption 3](#) gives $\mathbb{E}|\varepsilon_t|^\gamma < \infty$ for every $\gamma \in (0, \alpha)$, hence $\mathbb{E}[\log^+ |\varepsilon_t|] < \infty$, with $\log^+ x = \max(\log x, 0)$. The autoregression therefore admits a unique strictly stationary solution ([Rosenblatt, 2000](#); [Brockwell and Lindner, 2010](#)). Expressed in terms of (ε_t) , it reads

$$X_t = \sum_{j=-\infty}^{\infty} c_j \varepsilon_{t-j}, \quad (2.2)$$

where, by the geometric decay of the c_j and the log-moment condition, the series converges almost surely. Throughout, $(X_t)_{t \in \mathbb{Z}}$ denotes this solution. The representation (2.2) is two-sided, X_t depends on both past and future errors. [Figure 1](#) shows the response of X_t to a single error ε_j . In a noncausal autoregression the error is anticipated, X_t grows geometrically before date j and collapses right after it, whereas in a causal autoregression X_t jumps at date j and then decays geometrically. A MAR process combines the two patterns, a bubble that builds up at a rate set by the noncausal roots and then declines at a rate set by the causal roots.

[Assumption 3](#) holds for Student and Cauchy errors, and for α -stable errors whose skewness parameter lies in $(-1, 1)$ ([Samorodnitsky and Taqqu, 1994](#)), but excludes Gaussian errors, for which the model is not identified. We work with tails of Pareto type, rather than the weaker conditions of [Remark 1](#), because they are the framework of the literature on the prediction of MAR processes during explosive episodes ([Gouriéroux and Zakoïan, 2017](#); [Fries, 2022](#); [Gouriéroux et al., 2025](#); [de Truchis et al., 2025](#)).

2.2 Nonlinear state space representation

Since X_t depends on future errors, (2.2) does not describe how the process evolves from its past, which is what the Markov chain methods require. We therefore rewrite (X_t) in calendar time as a nonlinear state space model, driven by an *i.i.d.* innovation that is independent of the past of the process, as are the errors of a causal autoregression.

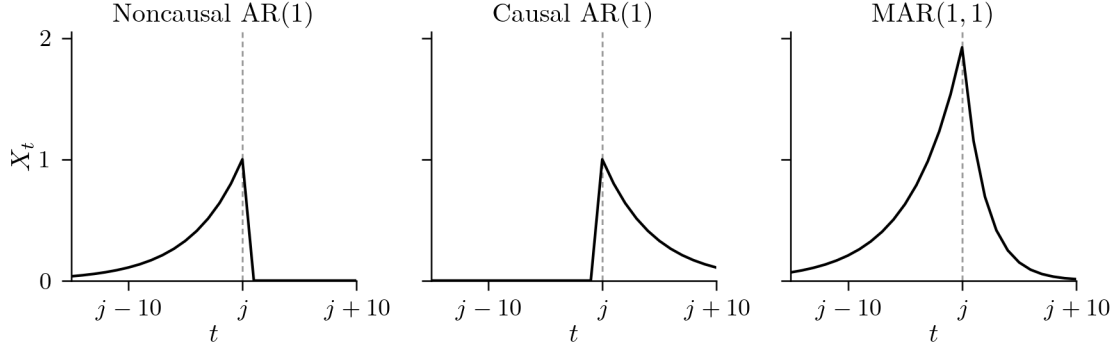


Figure 1: Response of X_t to a single error $\varepsilon_j = 1$, all other errors being zero. Left: noncausal AR(1), $(1 - 0.8F)X_t = \varepsilon_t$. Middle: causal AR(1), $(1 - 0.8B)X_t = \varepsilon_t$. Right: MAR(1, 1), $(1 - 0.8F)(1 - 0.6B)X_t = \varepsilon_t$. The dashed line marks the date j of the error.

The rest of this section follows the nonlinear state space representation of [Gouriéroux and Jasiak \(2026\)](#). We first write the model in companion form and separate its causal and noncausal components, then derive the predictive density of X_t given its past, from which we obtain the causal innovation and the transition kernel of the resulting Markov chain.

Following [Lanne and Saikkonen \(2011\)](#), let

$$\psi^*(z) = -\psi_s^{-1}z^s\psi(z^{-1}) = 1 - \psi_1^*z - \dots - \psi_s^*z^s,$$

a polynomial of degree s whose roots are the reciprocals of the roots of ψ , hence lie strictly inside the unit circle by [Assumption 1](#). Applying $-\psi_s^{-1}B^s$ to (2.1) shows that it is equivalent to

$$\theta(B)X_t = \eta_t, \quad \theta(z) = \psi^*(z)\phi(z) = 1 - \theta_1z - \dots - \theta_pz^p, \quad \eta_t = -\psi_s^{-1}\varepsilon_{t-s}, \quad (2.3)$$

with $p = r + s$, and we call η_t the shock at date t . The polynomial θ has r roots outside the unit circle (those of ϕ) and s roots inside (those of ψ^*), so (2.3) is a mixed causal–noncausal autoregression of order p with r causal and s noncausal roots.

The state vector $\mathbf{X}_t = (X_t, X_{t-1}, \dots, X_{t-p+1})' \in \mathbb{R}^p$ satisfies the companion form

$$\mathbf{X}_t = \Theta \mathbf{X}_{t-1} + \mathbf{e}_1 \eta_t, \quad \Theta = \begin{pmatrix} \theta_1 & \theta_2 & \dots & \theta_{p-1} & \theta_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \in \mathbb{R}^{p \times p}, \quad (2.4)$$

where $\mathbf{e}_1$ is the first vector of the canonical basis of $\mathbb{R}^p$. The eigenvalues of Θ are the reciprocals of the roots of θ , so that r of them have modulus strictly less than one (the causal eigenvalues) and s have modulus strictly larger than one (the noncausal

eigenvalues). Let

$$\Theta = A \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix} A^{-1}, \quad A^{-1} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix},$$

be a real Jordan decomposition, where $J_1 \in \mathbb{R}^{r \times r}$ and $J_2 \in \mathbb{R}^{s \times s}$ collect the causal and noncausal eigenvalues, respectively, and $A_1 \in \mathbb{R}^{r \times p}$, $A_2 \in \mathbb{R}^{s \times p}$. The noncausal eigenvalues of Θ are the reciprocals of the roots of ψ^* , that is, the roots of ψ , so that $\lambda_* = \min\{|\lambda| : \lambda \text{ eigenvalue of } J_2\}$. The causal and noncausal states $Z_{1,t} = A_1 \mathbf{X}_t$ and $Z_{2,t} = A_2 \mathbf{X}_t$ satisfy the state equations

$$Z_{1,t} = J_1 Z_{1,t-1} + a_1 \eta_t, \quad Z_{2,t} = J_2 Z_{2,t-1} + a_2 \eta_t, \quad a_i = A_i \mathbf{e}_1. \quad (2.5)$$

The observed state is recovered by the deterministic measurement equation $\mathbf{X}_t = AZ_t$, with $Z_t = (Z'_{1,t}, Z'_{2,t})'$, so that $(\mathbf{X}_t)$ and (Z_t) generate the same filtration. Solving (2.5) backward for $Z_{1,t}$ and forward for $Z_{2,t}$ gives

$$Z_{1,t} = \sum_{j=0}^{\infty} J_1^j a_1 \eta_{t-j}, \quad Z_{2,t} = - \sum_{j=0}^{\infty} J_2^{-(j+1)} a_2 \eta_{t+j+1} = \sum_{j \geq 1} d_j \eta_{t+j}, \quad (2.6)$$

where $d_j = -J_2^{-j} a_2 \in \mathbb{R}^s$ is the loading of the j -th future shock on the noncausal state.

By (2.6), the causal state $Z_{1,t}$ depends on current and past shocks, whereas the noncausal state $Z_{2,t}$ depends on future shocks only. Since $Z_{2,t}$ is nevertheless a function of $X_t, \dots, X_{t-p+1}$, the past of the process carries information on its future shocks, and the conditional law of X_t given its past must account for it. The predictive density of X_t given its past is expressed through the stationary density of the noncausal state $Z_{2,t}$. The construction of Gouriéroux and Jasiak (2026) relies on this density being positive and continuous, but does not establish it.

Lemma 1. *Under Assumptions 1 to 3, $Z_{2,t}$ admits a density f_{Z_2} with respect to the Lebesgue measure Leb on $\mathbb{R}^s$ that is positive, continuous and bounded on $\mathbb{R}^s$. This continuous version is unique.*

Since $Z_{2,t}$ is stationary, f_{Z_2} does not depend on t . Let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$ and, for $x \in \mathbb{R}$ and $\mathbf{y} = (y_1, \dots, y_p)' \in \mathbb{R}^p$, let $(x, \mathbf{y}_{-p}) = (x, y_1, \dots, y_{p-1})' \in \mathbb{R}^p$ denote the state obtained by adding the new observation x and dropping the oldest one. Under Assumptions 1 to 3, the predictive density of X_t given $\mathbf{X}_{t-1} = \mathbf{y}$ is

$$f(x | \mathbf{y}) = \frac{f_{Z_2}(A_2(x, \mathbf{y}_{-p}))}{f_{Z_2}(A_2 \mathbf{y})} |\det J_2| f_\eta(x - \boldsymbol{\theta}' \mathbf{y}), \quad x \in \mathbb{R}, \mathbf{y} \in \mathbb{R}^p, \quad (2.7)$$

where f_η is the density of η_t , which is positive, continuous and bounded on $\mathbb{R}$ by Assumption 2 since $\eta_t = -\psi_s^{-1} \varepsilon_{t-s}$ is a rescaling of ε_{t-s} . This is the predictive density of Gouriéroux and Jasiak (2026).

The matrices A_2 and J_2 are not unique, but (2.7) does not depend on the choice of the real Jordan decomposition: $\det J_2$ is the product of the noncausal eigenvalues, and replacing A_2 by LA_2 with L invertible multiplies the numerator and the denominator by the same Jacobian.

Since f_{Z_2} is positive, the denominator in (2.7) never vanishes, and $f(x \mid \mathbf{y})$ is positive and jointly continuous in $(x, \mathbf{y}) \in \mathbb{R} \times \mathbb{R}^p$ by Lemma 1 and Assumption 2, and it integrates to one for every $\mathbf{y}$, not only for almost every $\mathbf{y}$ (Section A.1). Throughout, we use this version of the predictive density.

Neither ε_t nor η_t is a causal innovation, but one can be constructed by the probability integral transform. Let $F(\cdot \mid \mathbf{y})$ be the distribution function of $f(\cdot \mid \mathbf{y})$, a continuous and strictly increasing bijection from $\mathbb{R}$ onto $(0, 1)$. Since (X_t) is a Markov chain of order p (Fries and Zakoian, 2019), the variables $U_t = F(X_t \mid \mathbf{X}_{t-1})$ are *i.i.d.* $\mathcal{U}[0, 1]$, and U_t is independent of $(X_k)_{k \leq t-1}$ (Gouriéroux and Jasiak, 2026). Hence $\mathbf{X}_t = G(\mathbf{X}_{t-1}, U_t)$, with $G(\mathbf{x}, u) = (F^{-1}(u \mid \mathbf{x}), x_1, \dots, x_{p-1})'$ for $\mathbf{x} = (x_1, \dots, x_p)'$ and $u \in (0, 1)$. The first coordinate draws the new observation, and the others shift the state down by one lag.

The state vector $(\mathbf{X}_t)$ is thus a homogeneous Markov chain on the state space $\mathbb{R}^p$, equipped with its Borel σ -field $\mathcal{B}(\mathbb{R}^p)$. Its transition kernel is

$$P(\mathbf{x}, B) = \int_{\mathbb{R}} \mathbf{1}_B((x, \mathbf{x}_{-p})) f(x \mid \mathbf{x}) dx = \mathbb{P}(G(\mathbf{x}, U) \in B), \quad \mathbf{x} \in \mathbb{R}^p, B \in \mathcal{B}(\mathbb{R}^p),$$

with $U \sim \mathcal{U}[0, 1]$. Its n -step iterates are defined recursively by $P^1 = P$ and $P^n(\mathbf{x}, B) = \int_{\mathbb{R}^p} P(\mathbf{x}, d\mathbf{y}) P^{n-1}(\mathbf{y}, B)$. Following Meyn and Tweedie (2009), we write $\mathbb{P}_{\mathbf{x}}$ for the law of the chain started at $\mathbf{X}_0 = \mathbf{x}$ and $\mathbb{E}_{\mathbf{x}}$ for the corresponding expectation, so that $P^n(\mathbf{x}, B) = \mathbb{P}_{\mathbf{x}}(\mathbf{X}_n \in B)$. The stationary MAR process corresponds to the initial law π of $\mathbf{X}_0$, which is invariant, $\pi P = \pi$. Under $\mathbb{P}_{\mathbf{x}}$, we write $\eta_t = X_t - \boldsymbol{\theta}'\mathbf{X}_{t-1}$ for the shock of the chain at date t , and $Z_{i,t} = A_i\mathbf{X}_t$, $i = 1, 2$, for its causal and noncausal states, which satisfy the state equations (2.5) but not the series representation (2.6). When a statement involves only the law of the stationary process, we work at $t = 0$ and write $\eta_1, \eta_2, \dots$ for the future shocks, so that $Z_{2,0} = \sum_{j \geq 1} d_j \eta_j$. Section 3 studies the chain from an arbitrary initial state, and not only under its stationary law.

3 Geometric ergodicity and β -mixing of $\text{MAR}(r, s)$ processes

The Geometric Ergodic Theorem (Meyn and Tweedie, 2009) requires the chain $(\mathbf{X}_t)$ to be irreducible and aperiodic, and to satisfy a Foster–Lyapunov drift condition outside a petite set. Section 3.1 establishes irreducibility, the smallness of bounded sets and aperiodicity, and Section 3.2 the drift condition. Section 3.3 deduces geometric ergodicity and geometric β -mixing.

3.1 Irreducibility, small sets and aperiodicity

Intuitively, since ε_t has a density that is positive on $\mathbb{R}$, so has the conditional distribution of X_t given $\mathbf{X}_{t-1}$, and the chain $(\mathbf{X}_t)$ should be irreducible with respect to the Lebesgue measure on $\mathbb{R}^p$. One step of the chain, however, renews only the first coordinate of the state, and the state is made of entirely new observations only after p steps. For every $n \geq p$, the n -step transition $P^n(\mathbf{x}, \cdot)$ then admits a density with respect to Leb that is positive and jointly continuous in the initial and final states (Lemma A.1), and the three properties of Proposition 1 below follow from this fact.

We refer the reader to Meyn and Tweedie (2009) for the formal definitions of irreducibility, small sets and aperiodicity.

Proposition 1. *Under Assumptions 1 to 3:*

- (i) *the chain $(\mathbf{X}_t)$ is Leb-irreducible, where Leb is the Lebesgue measure on $\mathbb{R}^p$;*
- (ii) *every bounded set $C \in \mathcal{B}(\mathbb{R}^p)$ is a small set for $(\mathbf{X}_t)$;*
- (iii) *the chain $(\mathbf{X}_t)$ is aperiodic.*

3.2 Drift condition

The last requirement of the Geometric Ergodic Theorem is a Foster–Lyapunov drift condition, the geometric drift condition (V4) of Meyn and Tweedie (2009). There exist a function $V : \mathbb{R}^p \rightarrow [1, \infty)$, constants $\rho \in (0, 1)$ and $b < \infty$, and a petite set $C \in \mathcal{B}(\mathbb{R}^p)$ such that

$$\Delta V(\mathbf{x}) \leq -(1 - \rho)V(\mathbf{x}) + b\mathbf{1}_C(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^p. \quad (3.1)$$

Here, $PV(\mathbf{x}) = \int_{\mathbb{R}^p} P(\mathbf{x}, d\mathbf{y})V(\mathbf{y}) = \mathbb{E}_{\mathbf{x}}[V(\mathbf{X}_1)]$, and the drift operator $\Delta V = PV - V$ is the expected one-step change in V . Outside C , the expected value of V after one step is at most ρ times its current value, so the chain is pulled back toward C at a geometric rate, and by Proposition 1 it suffices that C be bounded. We establish this condition, stated in Theorem 1, for every MAR(r, s) process.

For a causal autoregression, $V(\mathbf{x}) = 1 + \|\mathbf{x}\|^\kappa$, with $\kappa \in (0, 1]$ and $\kappa < \alpha$, satisfies the drift condition because the deterministic part of one step is a contraction and the shock is independent of the current state. Here neither holds. The noncausal block of the companion matrix expands, and by the forward representation (2.6) of the noncausal state the shock driving the chain from $\mathbf{X}_0$ to $\mathbf{X}_1$ is itself a component of $\mathbf{X}_0$. The causal component alone, $Z_{1,t}$, raises no such difficulty. The eigenvalues of J_1 have modulus smaller than one, so J_1 is a contraction in a suitably chosen norm, as for a causal autoregression.

Lemma 2. *If $r \geq 1$, there exist a norm $\|\cdot\|_*$ on $\mathbb{R}^r$ and $\rho_1 \in (0, 1)$ such that $\|J_1 v\|_* \leq \rho_1 \|v\|_*$ for all $v \in \mathbb{R}^r$.*

It remains to handle the noncausal component, $Z_{2,t}$, and the function we use for it is best explained by the dynamics of explosive episodes. The following paragraphs construct this function, denoted W , heuristically. The approximations they use are informal, and [Section 4](#) outlines the rigorous argument.

By (2.6) a large value $Z_{2,0} = z$ is typically produced by a single large future shock η_j , so that $z \approx d_j \eta_j$; for the law of $Z_{2,0}$ this single big shock principle is made precise by [Hult and Samorodnitsky \(2008\)](#). One period later,

$$Z_{2,1} = \sum_{k \geq 1} d_k \eta_{k+1} \approx d_{j-1} \eta_j = J_2 d_j \eta_j \approx J_2 z \quad \text{if } j \geq 2,$$

whereas $Z_{2,1}$ is moderate if $j = 1$, since η_1 does not appear in $Z_{2,1}$ at all. When $s = 1$ the noncausal state is scalar, so z carries no information on j , which remains random given $Z_{2,0} = z$. For the noncausal AR(1) with α -stable errors, the episode thus survives (asymptotically) each period with probability $\lambda_*^{-\alpha}$, and its lifetime is geometric ([Fries, 2022](#)). When $s \geq 2$, the direction of z reveals j , and the conditional law of j given z degenerates as $\|z\| \rightarrow \infty$.¹ Both regimes appear in the literature on the prediction of MAR processes during explosive episodes. [Gouriéroux et al. \(2025\)](#) show that, conditionally on a large $|X_t|$, the normalized future path converges to the tail process $\pm(c_{N+h}/c_N)_h$, where the random distance N to the large shock satisfies $\mathbb{P}(N = j) \propto |c_j|^\alpha$, so that only a countable set of patterns is charged. [de Truchis et al. \(2025\)](#) condition instead on a piece of the past trajectory and show that its shape identifies the pattern: the crash date remains random for the noncausal AR(1), but two consecutive observations determine the future path of the noncausal AR(2). However, these results describe the process conditionally on an extreme (asymptotic) level, whereas the drift condition must hold at every state.

Rather than exhibit the noncausal part of V directly, we list what it must satisfy. It must (i) dominate $\|z\|^\kappa$, so that V grows at infinity and the set C is bounded;² (ii) contract in one step; and (iii) be π -integrable, as the constant of [Theorem 3](#) requires. By the previous paragraph, when $s \geq 2$ requirement (ii) rules out functions of the size of z alone, such as $\|z\|^\kappa$. On the line $\mathbb{R}d_j$, $j \geq 2$, the state moves approximately to $J_2 z$, which is larger, so a contracting function must increase with the horizon j revealed by the direction of z . This horizon can be recorded through the conditional law of the individual future shocks given $Z_{2,0} = z$, which leads to functions of the form $\sum_{k \geq 1} w_k \mathbb{E}[|\eta_k|^\kappa \mid Z_{2,0} = z]$. The dynamics then determine the weights. After one period the k -th future shock has become the $(k-1)$ -th³, so the k -th term is carried onto the $(k-1)$ -th, and contraction at a rate $\rho < 1$ requires $w_k \leq \rho w_{k+1}$ for every k . The weights

¹This requires the loadings d_j to be pairwise non-collinear, that is, no power of J_2 to be a multiple of the identity.

²That is, $\|z\|^\kappa \leq C_W(1 + W(z))$ for some constant $C_W < \infty$ and all $z \in \mathbb{R}^s$.

³[Lemma 3](#) makes this exact at every state.

must grow at least geometrically. Taking $w_k = \omega^{-k}$ with $\omega \in (0, 1)$ gives contraction at rate exactly ω . A shock is thus weighted by the time remaining before it is realized, and the contraction comes from the passage of time.

With these weights this yields $W_0(z) = \mathbb{E}[\sum_{k \geq 1} \omega^{-k} |\eta_k|^\kappa \mid Z_{2,0} = z]$, which satisfies (i) and, formally,

$$\mathbb{E}_{\mathbf{x}} W_0(Z_{2,1}) = \omega W_0(z) - \mathbb{E}[|\eta_1|^\kappa \mid Z_{2,0} = z], \quad z = A_2 \mathbf{x},$$

a contraction at rate ω valid for every s . Nonetheless, it fails (iii). By the law of iterated expectations, its integral against f_{Z_2} equals $\mathbb{E}|\eta_t|^\kappa \sum_k \omega^{-k} = \infty$. The divergence comes from the moderate shocks, which are counted with weights ω^{-k} exploding with k although they hardly move the state. The k -th shock contributes $d_k \eta_k$ to $Z_{2,0}$, and since $\|d_k\|$ decays geometrically at a rate close to λ_*^{-1} , this contribution stays bounded unless $|\eta_k|$ is of order λ_*^k . We therefore count only the part of η_k above a threshold λ_0^k growing geometrically with k , with $\lambda_0 < \lambda_*$. Replacing $|u|^\kappa$ by $h_k(u) = (|u| - \lambda_0^k)_+^\kappa$ gives, for $\kappa \in (0, 1]$, $\lambda_0 > 1$ and $\omega \in (0, 1)$,

$$W(z) = \sum_{k \geq 1} \omega^{-k} \mathbb{E}[(|\eta_k| - \lambda_0^k)_+^\kappa \mid Z_{2,0} = z], \quad z \in \mathbb{R}^s. \quad (3.2)$$

The conditional expectations in (3.2) are taken in a version defined at every $z \in \mathbb{R}^s$, constructed in [Section 4.1](#).

Thresholding restores (iii) at the expense of (i), and the tension between the two determines the admissible parameters. Raising the threshold discards more of the state, while lowering it costs integrability. Both requirements hold when $1 < \lambda_0 < \lambda_*$ and $\lambda_0^{-(\alpha-\kappa)} < \omega < 1$, conditions that can be met precisely when $\kappa < \alpha$.

Our Lyapunov function combines the two components, the adapted norm of [Lemma 2](#) for the causal state, and W for the noncausal state. [Theorem 1](#) shows that it satisfies the drift condition (3.1).

Theorem 1. *Suppose [Assumptions 1 to 3](#) hold. Let*

$$\kappa \in (0, 1], \quad \kappa < \alpha, \quad \lambda_0 \in (1, \lambda_*), \quad \omega \in (\lambda_0^{-(\alpha-\kappa)}, 1), \quad (3.3)$$

and set

$$K = \begin{cases} \max(1, \|a_1\|_*^\kappa), & r \geq 1, \\ 1, & r = 0. \end{cases}$$

Then

$$V(\mathbf{x}) = \begin{cases} 1 + \|A_1 \mathbf{x}\|_*^\kappa + K W(A_2 \mathbf{x}), & r \geq 1, \\ 1 + W(A_2 \mathbf{x}), & r = 0, \end{cases} \quad \mathbf{x} \in \mathbb{R}^p,$$

with W given by (3.2), is finite and continuous on $\mathbb{R}^p$, $V(\mathbf{x}) \rightarrow \infty$ as $\|\mathbf{x}\| \rightarrow \infty$, and there exists $\rho_0 \in (0, 1)$ such that, for every $\rho \in (\rho_0, 1)$ and every bounded set $C_0 \in \mathcal{B}(\mathbb{R}^p)$, there exist $b < \infty$ and a bounded set $C \in \mathcal{B}(\mathbb{R}^p)$ containing C_0 such that

$$\Delta V(\mathbf{x}) \leq -(1 - \rho) V(\mathbf{x}) + b \mathbf{1}_C(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^p. \quad (3.4)$$

Moreover, $\int_{\mathbb{R}^p} V d\pi < \infty$.

Since every $\kappa \in (0, 1]$ with $\kappa < \alpha$ is admissible, Theorem 1 holds for every tail index $\alpha > 0$, including infinite variance ($\alpha < 2$) and infinite mean ($\alpha \leq 1$), whatever the noncausal roots and the balance of the tails. The rate ρ_0 is made explicit in Remark 2.

3.3 Geometric ergodicity and β -mixing

A chain is geometrically ergodic when the law of $\mathbf{X}_n$ converges to the invariant law at a rate that is geometric in n . By Proposition 1, the bounded set C of Theorem 1 is small, hence petite, and the Geometric Ergodic Theorem applies: $(\mathbf{X}_t)$ is positive recurrent and V -geometrically ergodic, and its unique invariant probability is the law π of $\mathbf{X}_0$.

Theorem 2. Under Assumptions 1 to 3, let V be as in Theorem 1 and π the law of $\mathbf{X}_0$. There exist $\vartheta > 1$ and $c_V < \infty$ such that

$$\sum_{n \geq 0} \vartheta^n \|P^n(\mathbf{x}, \cdot) - \pi\|_V \leq c_V V(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^p, \quad (3.5)$$

where $\|\nu\|_V = \sup\{|\int f d\nu| : f \text{ measurable, } |f| \leq V\}$ for a signed measure ν on $\mathcal{B}(\mathbb{R}^p)$.

We use the coefficient of absolute regularity of Volkonskii and Rozanov (1959). For two sub- σ -fields $\mathcal{F}$ and $\mathcal{G}$ of $\mathcal{A}$,

$$\beta(\mathcal{F}, \mathcal{G}) = \frac{1}{2} \sup \sum_{i=1}^I \sum_{j=1}^J |\mathbb{P}(F_i \cap G_j) - \mathbb{P}(F_i) \mathbb{P}(G_j)|,$$

the supremum being taken over all finite partitions $(F_i)_{i \leq I} \subseteq \mathcal{F}$ and $(G_j)_{j \leq J} \subseteq \mathcal{G}$ of Ω . The coefficient is nondecreasing in $\mathcal{F}$ and in $\mathcal{G}$ (Bradley, 2005). For a strictly stationary process $(\xi_t)_{t \in \mathbb{Z}}$, the coefficient $\beta_\xi(n) = \beta(\sigma(\xi_t : t \leq 0), \sigma(\xi_t : t \geq n))$, $n \geq 1$, measures how far the past and the future are from being independent when n periods separate them. The process is β -mixing, or absolutely regular, if $\beta_\xi(n) \rightarrow 0$, and geometrically β -mixing if $\beta_\xi(n) \leq c_\beta \varsigma^n$ for some $c_\beta < \infty$ and $\varsigma \in (0, 1)$.

A geometrically ergodic stationary Markov chain is geometrically β -mixing (Bradley, 2005). Theorem 3 makes this implication quantitative for the MAR(r, s) process, and transfers it from the state $(\mathbf{X}_t)$ to the observed process (X_t) .

Theorem 3. Under [Assumptions 1 to 3](#), with π , V , ϑ and c_V as in [Theorem 2](#), the chain $(\mathbf{X}_t)$ and the process (X_t) satisfy

$$\beta_X(n) \leq \beta_{\mathbf{X}}(n) \leq \frac{c_V}{2} \left(\int_{\mathbb{R}^p} V d\pi \right) \vartheta^{-n}, \quad n \geq 1. \quad (3.6)$$

More precisely,

$$\sum_{n \geq 1} \vartheta^n \beta_{\mathbf{X}}(n) \leq \frac{c_V}{2} \int_{\mathbb{R}^p} V d\pi < \infty. \quad (3.7)$$

In particular $(\mathbf{X}_t)$ and (X_t) are geometrically β -mixing with rate $\varsigma = \vartheta^{-1} \in (0, 1)$, hence also strongly mixing at that rate.

Let ρ , b and C be the constants of [\(3.4\)](#). Since π is invariant, $\int PV d\pi = \int V d\pi < \infty$, and integrating [\(3.4\)](#) against π gives $\int V d\pi \leq b\pi(C)/(1 - \rho)$, hence

$$\beta_X(n) \leq \beta_{\mathbf{X}}(n) \leq \frac{c_V b}{2(1 - \rho)} \vartheta^{-n}, \quad n \geq 1.$$

The dependence between the past and the future of a $\text{MAR}(r, s)$ process therefore vanishes at a geometric rate, with a constant controlled by the drift condition.

4 Outline of the proof of [Theorem 1](#)

This section outlines the proof of [Theorem 1](#). [Section 4.1](#) constructs the conditional laws of the future shocks and establishes the shift identity. [Section 4.2](#) states the two propositions on W , and combines them with [Lemma 2](#) into a one-step bound on PV , while [Remark 1](#) identifies the properties of the error law used by the proof.

4.1 The conditional law of the future shocks

Since W involves every future shock, we need the conditional law of each η_k given the noncausal state, defined at every z rather than almost everywhere, because [\(3.4\)](#) must hold at every state. Removing the k -th shock from the noncausal state gives

$$Z_{2,0} = d_k \eta_k + R_k, \quad R_k = \sum_{j \geq 1, j \neq k} d_j \eta_j,$$

with R_k independent of η_k . By [Lemma A.4](#), R_k has a positive, continuous and bounded density f_{R_k} , with bounds uniform in k , which makes the following construction possible.

For $k \geq 1$ and $z \in \mathbb{R}^s$, let

$$p_k(u \mid z) = \frac{f_{\eta}(u) f_{R_k}(z - d_k u)}{f_{Z_2}(z)}, \quad u \in \mathbb{R}, \quad (4.1)$$

and, for measurable $h : \mathbb{R} \rightarrow [0, \infty]$,

$$\mathbb{E}[h(\eta_k) \mid Z_{2,0} = z] = \int_{\mathbb{R}} h(u) p_k(u \mid z) du \in [0, \infty]. \quad (4.2)$$

Conditional expectations and probabilities given $Z_{2,0} = z$ are always taken in this version, which is defined at every $z \in \mathbb{R}^s$. By [Section A.3.4](#), $p_k(\cdot \mid z)$ is indeed the conditional density of η_k given $Z_{2,0} = z$. For all measurable $h \geq 0$ and Borel $B \subseteq \mathbb{R}^s$,

$$\mathbb{E}[h(\eta_k) \mathbf{1}\{Z_{2,0} \in B\}] = \int_B \mathbb{E}[h(\eta_k) \mid Z_{2,0} = z] f_{Z_2}(z) dz. \quad (4.3)$$

For $k = 1$, this density simplifies to

$$p_1(u \mid z) = |\det J_2| \frac{f_{Z_2}(J_2 z + a_2 u)}{f_{Z_2}(z)} f_\eta(u), \quad u \in \mathbb{R}, \quad (4.4)$$

which by [\(2.7\)](#) is the conditional density of the shock η_1 of the chain given $\mathbf{X}_0 = \mathbf{x}$, for any $\mathbf{x}$ with $A_2 \mathbf{x} = z$.

Lemma 3. *Let $k \geq 1$ and let $h : \mathbb{R} \rightarrow [0, \infty]$ be measurable. For every $\mathbf{x} \in \mathbb{R}^p$,*

$$\mathbb{E}_{\mathbf{x}}[m(Z_{2,1})] = \mathbb{E}[h(\eta_{k+1}) \mid Z_{2,0} = A_2 \mathbf{x}], \quad m(z) = \mathbb{E}[h(\eta_k) \mid Z_{2,0} = z].$$

For the stationary process the identity reduces to the law of iterated expectations: $Z_{2,1} = \sum_{j \geq 1} d_j \eta_{j+1}$, so that $m(Z_{2,1}) = \mathbb{E}[h(\eta_{k+1}) \mid Z_{2,1}]$, and $Z_{2,0} = J_2^{-1}(Z_{2,1} - a_2 \eta_1)$ with η_1 independent of $(Z_{2,1}, \eta_{k+1})$. The causal state $Z_{1,0}$, a function of past shocks, carries no further information on the future ones. What the lemma adds, and what [\(3.4\)](#) requires, is that it holds at every initial state with no error term.

4.2 Drift of the noncausal component

Throughout the rest of this subsection the constants satisfy [\(3.3\)](#). We also fix a moment order $\gamma \in (\kappa, \alpha)$ such that $q = \frac{\lambda_0^{-(\gamma-\kappa)}}{\omega} < 1$, which is possible since $q < 1$ if and only if $\gamma > \kappa + \log(1/\omega)/\log \lambda_0$, and this lower bound is smaller than α precisely because $\omega > \lambda_0^{-(\alpha-\kappa)}$. Then, writing η_t for a generic shock, $\mathbb{E}|\eta_t|^\gamma < \infty$ and $\omega^{-k} \mathbb{E}[h_k(\eta_t)] \leq q^k \mathbb{E}|\eta_t|^\gamma$ for every $k \geq 1$, so that q is the geometric rate at which the terms of W decay. The two propositions below rest on the decomposition $Z_{2,0} = d_k \eta_k + R_k$ and on two properties of the density f_{R_k} of the remaining shocks, uniform in k : boundedness and a local ratio bound.

Proposition 2. *Suppose [Assumptions 1 to 3](#) hold, and let the constants satisfy [\(3.3\)](#). Then W is finite and continuous on $\mathbb{R}^s$, and there exists a constant $C_W < \infty$ such that*

$$\|z\|^\kappa \leq C_W(1 + W(z)), \quad z \in \mathbb{R}^s, \quad (4.5)$$

and

$$\int_{\mathbb{R}^s} W(z) f_{Z_2}(z) dz = \sum_{k \geq 1} \omega^{-k} \mathbb{E}[h_k(\eta_t)] < \infty. \quad (4.6)$$

The bound (4.5) is the domination requirement, and (4.6) is integrability.

Proposition 3. *Suppose Assumptions 1 to 3 hold, let the constants satisfy (3.3), and let $\rho_2 \in (\omega, 1)$. There exists $b_2 < \infty$ such that, for every $\mathbf{x} \in \mathbb{R}^p$ with $z = A_2 \mathbf{x}$,*

$$\mathbb{E}_{\mathbf{x}} W(Z_{2,1}) \leq \rho_2 W(z) - \mathbb{E}[h_1(\eta_1) \mid Z_{2,0} = z] + b_2. \quad (4.7)$$

By Lemma 3, after one period each future shock is one period closer and its weight gains the factor ω , which alone would give contraction at rate ω . Its threshold drops at the same time, from λ_0^{k+1} to λ_0^k , and this drop costs a factor close to one, which turns ω into ρ_2 , and the constant b_2 , which collects the moderate shocks. The inequality holds at every z , so compact sets need no separate treatment.

Propositions 2 and 3 and Lemma 2 combine into a one-step bound on PV . Suppose first $r \geq 1$, and let $\mathbf{x} \in \mathbb{R}^p$, $z_1 = A_1 \mathbf{x}$ and $z = A_2 \mathbf{x}$. By (2.5) the two states move as $Z_{1,1} = J_1 z_1 + a_1 \eta_1$ and $Z_{2,1} = J_2 z + a_2 \eta_1$, and, writing $m_1(z) = \mathbb{E}[h_1(\eta_1) \mid Z_{2,0} = z]$, the proof of Theorem 1 establishes

$$PV(\mathbf{x}) - 1 \leq \underbrace{\rho_1^\kappa \|z_1\|_*^\kappa}_{(a)} + \underbrace{K \rho_2 W(z)}_{(b)} + \underbrace{(\|a_1\|_*^\kappa - K) m_1(z)}_{(c)} + \underbrace{\|a_1\|_*^\kappa \lambda_0^\kappa + K b_2}_{(d)}, \quad (4.8)$$

where $\rho_2 \in (\omega, 1)$ and $b_2 < \infty$ are the constants of Proposition 3. Term (a) is the contraction of J_1 in the adapted norm and (b) is Proposition 3, multiplied by the weight K of W in V . Term (c) is the only one that can grow with z . It accounts for the shock η_1 , which leaves the noncausal state, where W carried it with weight ω^{-1} , and enters the causal one through $a_1 \eta_1$. It is nonpositive as soon as $K \geq \|a_1\|_*^\kappa$. Term (d) is constant and is absorbed into a bounded set C . If $r = 0$ the causal terms are absent, $K = 1$, and (c) is $-m_1(z) \leq 0$.

Remark 1. *The proof of Theorem 1 uses Assumption 3 only through two of its consequences: the moment condition $\mathbb{E}|\eta_t|^\gamma < \infty$ for $\gamma \in (0, \alpha)$, and the local ratio bound*

$$H_\eta(\delta) = \sup_{|u-v| \leq \delta} \frac{f_\eta(u)}{f_\eta(v)} < \infty \quad \text{for every } \delta \geq 0,$$

which bounds how much f_η can vary over intervals of fixed length. Furthermore, the results of Section 2 and Proposition 1 use it only through the log-moment $\mathbb{E}[\log^+ |\varepsilon_t|] < \infty$, and the regular variation of the tails enters only the proof of the local ratio bound (Lemma A.2). Consequently, Theorems 1 to 3 remain valid when Assumption 3 is replaced by these two conditions, with α replaced by some $\alpha_0 > 0$, both in the moment condition and in (3.3). These conditions do not require tails of Pareto type. They hold, for instance, for the Laplace and logistic laws. The local ratio bound fails for the Gaussian law.

5 Implications

This section draws the consequences of [Theorems 1 to 3](#) for explosive episodes, inference and prediction. Throughout, [Assumptions 1 to 3](#) hold and $\rho \in (\rho_0, 1)$ is fixed, with b and C as in [Theorem 1](#).

5.1 Duration of explosive episodes

The drift condition [\(3.4\)](#) states that, outside the bounded set C , V is expected to shrink by the factor ρ at each date. Since C is bounded and contains a ball centered at the origin ([Theorem 1](#) allows this choice), it can be read as the normal regime of the process, and the time needed to reach it as the duration of an explosive episode. Let $\tau_C = \inf\{t \geq 1 : \mathbf{X}_t \in C\}$ be the return time to C .

Proposition 4. *For every $\mathbf{x} \in \mathbb{R}^p$ and $n \geq 0$,*

$$\mathbb{P}_{\mathbf{x}}(\tau_C > n) \leq (\rho V(\mathbf{x}) + b)\rho^n, \quad \mathbb{E}_{\mathbf{x}}[\tau_C] \leq \frac{\log(V(\mathbf{x}) + b/\rho)}{\log(1/\rho)}. \quad (5.1)$$

[Proposition 4](#) gives quantitative content to the informal statement that the bubbles of a MAR process eventually burst. Starting from a state $\mathbf{x}$ inside an explosive episode, the time needed to return to the normal regime has geometric tails with rate ρ , with a constant proportional to $V(\mathbf{x})$. The expected duration grows at most like $\log V(\mathbf{x})$.

The lifetime of a bubble is known to be geometric conditionally on an extreme level ([Fries, 2022](#)). [Proposition 4](#) holds instead at every state, at the price of the factor $\log V(\mathbf{x})$. A drift condition does not capture the exact dependence on the starting point, so our bound grows with the scale of the bubble while the limiting lifetime does not.

Remark 2. *The proof of [Theorem 1](#) gives $PV \leq \rho_0 V + b_0$ for some $b_0 < \infty$, with $\rho_0 = \max(\rho_1^\kappa, \rho_2)$, where ρ_1 and ρ_2 are the rates of [Lemma 2](#) and [Proposition 3](#), and $\rho_0 = \rho_2$ if $r = 0$. Since ρ_1 can be taken arbitrarily close to the spectral radius λ_ϕ^{-1} of J_1 , and ρ_2 to $\lambda_*^{-(\alpha-\kappa)}$ by letting $\lambda_0 \rightarrow \lambda_*$ and $\omega \rightarrow \lambda_0^{-(\alpha-\kappa)}$, the drift condition [\(3.4\)](#) holds, for suitable constants, for every $\rho \in (\max(\lambda_\phi^{-\kappa}, \lambda_*^{-(\alpha-\kappa)}), 1)$, with $\lambda_\phi^{-\kappa} = 0$ if $r = 0$, while the construction never gives a rate below $\lambda_*^{-(\alpha-\kappa)}$. The first term is the deflation of the causal component after an explosive episode, and the second is the survival probability $\lambda_*^{-\alpha}$ of the episode times its growth λ_*^κ . For the noncausal AR(1) it tends, as $\kappa \rightarrow 0$, to the per-period survival probability of the bubbles in [Fries \(2022\)](#). When $\alpha \leq 1$ and $r \geq 1$, the two terms are equal for $\kappa = \alpha \log \lambda_*/(\log \lambda_\phi + \log \lambda_*)$, and the drift condition then holds for every $\rho \in (e^{-1/T}, 1)$, with $T = \alpha^{-1}(1/\log \lambda_\phi + 1/\log \lambda_*)$. By [Proposition 4](#), the durations of explosive episodes then have geometric tails with any rate $\rho > e^{-1/T}$, that is, with a time constant arbitrarily close to T , the sum of the time constants $1/(\alpha \log \lambda_\phi)$ and $1/(\alpha \log \lambda_*)$ of the causal and noncausal parts.*

5.2 Consequences for inference

Many inference procedures for dependent data only require strong mixing, such as the central limit theorem of [Ibragimov \(1962\)](#), heteroskedasticity and autocorrelation consistent estimators ([Newey and West, 1987](#); [Andrews, 1991](#)), block bootstraps ([Künsch, 1989](#)) and subsampling ([Politis and Romano, 1994](#)). For MAR processes, these procedures were already available through the strong mixing of linear processes ([Doukhan, 1994](#)).

The β -mixing established in [Theorem 3](#) is stronger, as it allows blocking arguments that replace the data by independent blocks ([Berbee, 1979](#)). The most natural application is to extremes. The explosive episodes of MAR processes produce clusters of extremes ([Fries and Zakoïan, 2019](#)), and the theory of the Hill estimator ([Drees, 2000](#)) and of cluster functionals such as the extremal index ([Drees and Rootzén, 2010](#)) is built precisely on β -mixing and independent blocks. Blocking also gives functional central limit theorems for empirical processes under bracketing entropy conditions ([Doukhan et al., 1995](#)). For simple classes, such as the empirical distribution function or smooth parametric families, strong mixing at a geometric rate already suffices ([Andrews and Pollard, 1994](#); [Rio, 2017](#)), so the gain from β -mixing lies in larger classes of functions.

5.3 Prediction and empirical risk minimization

Forecasting MAR processes is difficult, since their predictive density is nonlinear and involves the stationary density f_{Z_2} of the noncausal state. Rather than estimating this density, one can learn a predictor of X_{t+h} directly, by empirical risk minimization (ERM) over a class of prediction rules. The theory of ERM for dependent data controls the gap between empirical and population risks through blocking arguments ([Yu, 1994](#)), at a cost governed by the β -mixing coefficients.

Since β_X decays geometrically ([Theorem 3](#)), blocks of length of order $\log n$ make the cost of blocking negligible ([Yu, 1994](#)), so bounds derived for independent data apply to MAR data with an effective sample size of order $n/\log n$, as in the analysis of finite-lag ERM by [Brownlees and Guðmundsson \(2025\)](#). Our results open the way to a convergence proof for the approach of [Dumitrescu et al. \(2026\)](#), who learn the predictive density of MAR processes with a mixture density network, that is, by empirical risk minimization with the negative log-likelihood as loss and finitely many lags as inputs. On stationary data, this blocking argument reduces oracle inequalities for this procedure, which bound the risk of the fitted network by the best risk in the class plus an estimation term, to their counterparts for independent data, under the same conditions on the loss.

6 Extensions to multivariate processes, exogenous regressors and moving average components

This section extends [Theorems 1 to 3](#) to mixed causal–noncausal VAR processes, and deduces geometric β -mixing for MAR models with exogenous regressors (MARX and MARDL) and for MAR and VAR models with moving average components (MARMA and VARMA).

6.1 Mixed causal–noncausal VAR processes

Let $(\eta_t)_{t \in \mathbb{Z}}$ be an *i.i.d.* sequence of random vectors in $\mathbb{R}^m$ and consider the VAR(p) equation

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \eta_t, \quad t \in \mathbb{Z}, \quad (6.1)$$

with $\Phi_j \in \mathbb{R}^{m \times m}$, where the roots of $\det(I_m - \Phi_1 z - \cdots - \Phi_p z^p)$ may lie on both sides of the unit circle, as in [Davis and Song \(2020\)](#) and [Gouriéroux and Jasiak \(2026\)](#). The state $\mathbf{Y}_t = (Y'_t, \dots, Y'_{t-p+1})' \in \mathbb{R}^{mp}$ satisfies $\mathbf{Y}_t = \Psi_c \mathbf{Y}_{t-1} + E \eta_t$, where Ψ_c is the block companion matrix of (6.1) and $E = (I_m, 0, \dots, 0)' \in \mathbb{R}^{mp \times m}$. Let n_1 and $n_2 = mp - n_1$ be the numbers of eigenvalues of Ψ_c , counted with multiplicity, of modulus smaller and larger than one. We keep the notation of [Section 2.2](#) for the corresponding objects: $\Psi_c = A \operatorname{diag}(J_1, J_2) A^{-1}$ with $J_i \in \mathbb{R}^{n_i \times n_i}$ and $A^{-1} = (A'_1, A'_2)'$, $a_i = A_i E \in \mathbb{R}^{n_i \times m}$, $Z_{i,t} = A_i \mathbf{Y}_t$, $d_j = -J_2^{-j} a_2 \in \mathbb{R}^{n_2 \times m}$, $R_k = Z_{2,0} - d_k \eta_k$, and λ_* is the smallest modulus of the eigenvalues of J_2 ; matrix norms are operator norms. For $m = 1$ and $\Phi_j = \theta_j$, (6.1) is (2.3), the state dimension mp is $p = r + s$, and these objects reduce to the univariate ones.

Assumption 4. *The VAR (6.1) satisfies:*

- (ι) Ψ_c has no eigenvalue on the unit circle, and $n_2 \geq 1$;
- (υ) η_t has a density g with respect to the Lebesgue measure on $\mathbb{R}^m$, positive and continuous;
- ($\iota\iota$) $H_g(\delta) = \sup_{\|u-v\| \leq \delta} g(u)/g(v) < \infty$ for every $\delta \geq 0$;
- ($\iota\nu$) $\mathbb{E} \|\eta_t\|^{\alpha_0} < \infty$ for some $\alpha_0 > 0$.

[Assumption 4](#) does not require Φ_p to be nonsingular, since zero eigenvalues of Ψ_c are causal, and ($\iota\iota$)–($\iota\nu$) are the conditions of [Remark 1](#). Condition ($\iota\iota$) implies that g is bounded, is preserved by invertible linear maps, holds for independent components satisfying [Assumptions 2 and 3](#), and fails for Gaussian laws.

The following theorem extends [Theorems 1 to 3](#) to the mixed causal–noncausal VAR.

Theorem 4. Suppose [Assumption 4](#) holds. Then [\(6.1\)](#) admits a unique strictly stationary solution (Y_t) , and $(\mathbf{Y}_t)$ is a Markov chain on $\mathbb{R}^{mp}$ with transition kernel P . Let κ , λ_0 and ω satisfy [\(3.3\)](#) with α_0 in place of α , let $\|\cdot\|_*$ be given by [Lemma 2](#) for J_1 , $K = \max(1, \|a_1\|_*^\kappa)$, where $\|a_1\|_* = \sup_{\|u\| \leq 1} \|a_1 u\|_*$, and

$$W(z) = \sum_{k \geq 1} \omega^{-k} \mathbb{E} \left[(\|\eta_k\| - \lambda_0^k)_+^\kappa \mid Z_{2,0} = z \right], \quad V(\mathbf{x}) = 1 + \|A_1 \mathbf{x}\|_*^\kappa + K W(A_2 \mathbf{x}),$$

with conditional expectations as in [\(4.1\)](#) and [\(4.2\)](#), g in place of f_η , and $K = 1$ and no causal term if $n_1 = 0$. The conclusions of [Proposition 1](#) and [Theorems 1](#) and [2](#) hold for $(\mathbf{Y}_t)$ and V , with $\mathbb{R}^{mp}$, $\mathbb{R}^{n_2}$ and $\mathbf{Y}_0$ in place of $\mathbb{R}^p$, $\mathbb{R}^s$ and $\mathbf{X}_0$, and, with π the law of $\mathbf{Y}_0$ and ϑ and c_V the constants of [Theorem 2](#) for $(\mathbf{Y}_t)$,

$$\beta_Y(n) \leq \beta_{\mathbf{Y}}(n) \leq \frac{c_V}{2} \left(\int_{\mathbb{R}^{mp}} V d\pi \right) \vartheta^{-n}, \quad n \geq 1.$$

For $m = 1$, [Theorem 4](#) recovers [Theorems 1](#) to [3](#) under the conditions of [Remark 1](#). The new arguments concern the existence and uniqueness of the stationary solution and the densities f_{Z_2} and f_{R_k} .

The noncausal VAR $\Pi(B)\Gamma(F)Y_t = \varepsilon_t$ of [Lanne and Saikkonen \(2013\)](#), with causal and noncausal polynomials of degrees r and s , is of the form [\(6.1\)](#) with $p = r + s$, $n_2 = ms$ and $\eta_t = -\Gamma_s^{-1} \varepsilon_{t-s}$ when Γ_s is nonsingular.

6.2 MAR models with exogenous regressors

The MARX model of [Hecq et al. \(2020\)](#) is

$$\psi(F)\phi(B)y_t = \beta'x_t + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (6.2)$$

with regressors $x_t \in \mathbb{R}^{m_x}$ and $\beta \in \mathbb{R}^{m_x}$. Unlike [Hecq et al. \(2020\)](#), who allow general two-sided linear regressors, we assume that the regressors follow a VAR $\Phi_x(B)x_t = \varepsilon_{x,t}$ of the form [\(6.1\)](#), with $\Phi_x(z) = I_{m_x} - \Phi_{x,1}z - \dots - \Phi_{x,p_x}z^{p_x}$ and possibly noncausal roots. This covers their VAR(2) in which x_t does not depend on y_t , provided its errors are independent, and their simulation designs with Student regressors. It excludes infinite-order linear regressors, which have no finite-dimensional Markov representation, and, by [Assumption 5](#) below, Gaussian regressor innovations. Applying $-\psi_s^{-1}B^s$ to [\(6.2\)](#) shows that $Y_t = (y_t, x'_t)'$ satisfies

$$\begin{pmatrix} \theta(B) & \psi_s^{-1}\beta'B^s \\ 0 & \Phi_x(B) \end{pmatrix} Y_t = \begin{pmatrix} \eta_t \\ \varepsilon_{x,t} \end{pmatrix}, \quad \eta_t = -\psi_s^{-1}\varepsilon_{t-s}, \quad (6.3)$$

a VAR of the form [\(6.1\)](#), of dimension $1 + m_x$ and order $\bar{p} = \max(p, p_x)$, with $p = r + s$ the order of θ , whose determinant $\theta(z) \det \Phi_x(z)$ has inside the unit circle the s roots of ψ^* and the roots of $\det \Phi_x$ that lie there, if any.

The mixed causal–noncausal autoregressive distributed lag (MARDL) model of [Hecq et al. \(2020\)](#) lets the regressors enter with leads and lags,

$$\psi(F)\phi(B)y_t = \sum_{i=0}^{k_x} \beta'_i x_{t+s-i} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (6.4)$$

with $\beta_i \in \mathbb{R}^{m_x}$. Applying $-\psi_s^{-1}B^s$ turns it again into a VAR in B for Y_t , of order $\max(p, p_x, k_x)$, after a normalization of the leading coefficient ([Section A.7](#)).

Assumption 5. ϕ and ψ satisfy [Assumption 1](#), the companion matrix of Φ_x has no eigenvalue on the unit circle, (ε_t) and $(\varepsilon_{x,t})$ are independent i.i.d. sequences, and the densities of η_t and $\varepsilon_{x,t}$ satisfy (ν) – (ν) of [Assumption 4](#), with a common α_0 .

For η_t , these conditions hold under [Assumptions 2](#) and [3](#).

Corollary 1. Under [Assumption 5](#), [Theorem 4](#) applies to the VAR representation of $(y_t, x'_t)'$, both for the MARX model (6.2) and for the MARDL model (6.4). In particular, $(y_t, x'_t)'$ is geometrically β -mixing.

[Corollary 1](#) gives the joint mixing of the dependent variable and the regressors, a property on which limit theorems for MARX estimators with dependent regressors can rest. On the other hand, its conditions are stronger than those of [Hecq et al. \(2020\)](#), which only require the errors to be mean independent of the regressors.

6.3 Models with moving average components

The MARMA(r, s, r', s') model of [Fries \(2022\)](#), [Hecq and Velasquez-Gaviria \(2025\)](#) and [Dumitrescu et al. \(2026\)](#) is

$$\psi(F)\phi(B)\tilde{X}_t = v(F)\zeta(B)\varepsilon_t, \quad t \in \mathbb{Z}, \quad (6.5)$$

with $\zeta(z) = 1 - \zeta_1 z - \dots - \zeta_{r'} z^{r'}$, $v(z) = 1 - v_1 z - \dots - v_{s'} z^{s'}$, and (ε_t) , ϕ and ψ as before. The mixed causal–noncausal vector autoregressive moving average (VARMA) equation is

$$\Phi(B)Y_t = \Lambda(B)\eta_t, \quad t \in \mathbb{Z}, \quad (6.6)$$

with $Y_t, \eta_t \in \mathbb{R}^m$, $\Phi(z) = I_m - \Phi_1 z - \dots - \Phi_p z^p$ and $\Lambda(z) = \Lambda_0 + \Lambda_1 z + \dots + \Lambda_q z^q$, $\Phi_j, \Lambda_j \in \mathbb{R}^{m \times m}$, where $\det \Phi$ has at least one root inside the unit circle and none on it.

Corollary 2. (a) Under [Assumptions 1](#) to [3](#), $\tilde{X}_t = v(F)\zeta(B)X_t$ is the unique strictly stationary solution of (6.5), and $\beta_{\tilde{X}}(n) \leq \beta_X(n - r' - s')$ for $n > r' + s'$, so that [Theorem 3](#) applies to $(\tilde{X}_t)$ with $n - r' - s'$ in place of n . (b) If η_t satisfies (ν) – (ν) of [Assumption 4](#), then $Y_t = \text{adj } \Phi(B)\Lambda(B)\mathcal{Y}_t$ is the unique strictly stationary solution of (6.6), where $(\mathcal{Y}_t)$ is the strictly stationary solution of the VAR $\det \Phi(B)\mathcal{Y}_t = \eta_t$, to which [Theorem 4](#) applies, and $\beta_Y(n) \leq \beta_{\mathcal{Y}}(n - k)$ for $n > k = \deg(\text{adj } \Phi \Lambda)$. In both cases, the solution is geometrically β -mixing.

No condition on the roots of ζ , v and $\det \Lambda$, nor on common roots with the autoregressive part, is needed. These conditions identify the model (Hecq and Velasquez-Gaviria, 2025) but play no role in mixing.

7 Conclusion

We have shown that the state vector of a $\text{MAR}(r, s)$ process with *i.i.d.* errors is V -geometrically ergodic under mild conditions on the error density, with a finite moment of arbitrarily small order, so that errors with infinite mean or infinite variance are allowed. The proof relies on a Lyapunov function that weights each future shock by the time remaining before it is realized, so that the drift condition comes from the passage of time. This implies that $\text{MAR}(r, s)$ processes are geometrically β -mixing, for any number of causal and noncausal roots, real or complex. The result extends to multivariate processes and to models with exogenous regressors or moving average components.

To the best of our knowledge, this is the first β -mixing result for linear processes with a two-sided moving average representation, for which only strong mixing was available (Doukhan, 1994; Rosenblatt, 2000). Beyond the procedures that only need strong mixing, β -mixing makes blocking arguments available. It provides the mixing part of the conditions under which the Hill estimator and cluster functionals are analyzed, and opens the way to oracle inequalities for nonparametric empirical risk minimization procedures that learn the predictive density of MAR processes, whose dependence on future shocks makes it hard to compute directly.

References

- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.
- Andrews, D. W. K. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review*, 62(1):119–132.
- Berbee, H. C. P. (1979). *Random Walks with Stationary Increments and Renewal Theory*, volume 112 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam.
- Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York, 2 edition.
- Brockwell, P. J. and Lindner, A. (2010). Strictly stationary solutions of autoregressive moving average equations. *Biometrika*, 97(3):765–772.

- Brockwell, P. J., Lindner, A., and Vollenbröker, B. (2012). Strictly stationary solutions of multivariate ARMA equations with i.i.d. noise. Annals of the Institute of Statistical Mathematics, 64(6):1089–1119.
- Brownlees, C. and Guðmundsson, G. S. (2025). Performance of empirical risk minimization for linear regression with dependent data. Econometric Theory, 41(2):391–420.
- Cavaliere, G., Nielsen, H. B., and Rahbek, A. (2020). Bootstrapping noncausal autoregressions: With applications to explosive bubble modeling. Journal of Business & Economic Statistics, 38(1):55–67.
- Cubadda, G., Giancaterini, F., and Hecq, A. (2023). Comparative analysis of multivariate mixed causal and noncausal process representations. Working paper, Chapter 5 of F. Giancaterini’s doctoral thesis, Maastricht University.
- Davis, R. A. and Song, L. (2020). Noncausal vector AR processes with application to economic time series. Journal of Econometrics, 216(1):246–267.
- Davydov, Y. A. (1973). Mixing conditions for Markov chains. Theory of Probability and Its Applications, 18(2):312–328.
- de Truchis, G., Fries, S., and Thomas, A. (2025). Forecasting extreme trajectories using seminorm representations. HAL working paper hal-05007564.
- de Truchis, G. and Thomas, A. (2026). Laurent series expansion for $MA(\infty)$ representation of mixed causal–noncausal autoregressive processes. Journal of Time Series Analysis.
- Doukhan, P. (1994). Mixing: Properties and Examples, volume 85 of Lecture Notes in Statistics. Springer, New York.
- Doukhan, P., Massart, P., and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. Annales de l’Institut Henri Poincaré, Probabilités et Statistiques, 31(2):393–427.
- Drees, H. (2000). Weighted approximations of tail processes for β -mixing random variables. The Annals of Applied Probability, 10(4):1274–1301.
- Drees, H. and Rootzén, H. (2010). Limit theorems for empirical processes of cluster functionals. The Annals of Statistics, 38(4):2145–2186.
- Dumitrescu, E., Peignon, J., and Thomas, A. (2026). Tail-aware density forecasting of locally explosive time series: A neural network approach. arXiv:2601.14049.
- Fries, S. (2022). Conditional moments of noncausal alpha-stable processes and the prediction of bubble crash odds. Journal of Business & Economic Statistics, 40(4):1596–1616.

- Fries, S. and Zakoïan, J.-M. (2019). Mixed causal-noncausal AR processes and the modelling of explosive bubbles. Econometric Theory, 35(6):1234–1270.
- Gouriéroux, C. and Jasiak, J. (2026). Nonlinear fore(back)casting and innovation filtering for causal–noncausal VAR models. Journal of Financial Econometrics, 24(2):nbag005.
- Gouriéroux, C., Lu, Y., and Robert, C. Y. (2025). The causal–noncausal tail processes. arXiv:2506.04046.
- Gouriéroux, C. and Zakoïan, J.-M. (2017). Local explosion modelling by non-causal process. Journal of the Royal Statistical Society Series B: Statistical Methodology, 79(3):737–756.
- Hecq, A., Issler, J. V., and Telg, S. (2020). Mixed causal–noncausal autoregressions with exogenous regressors. Journal of Applied Econometrics, 35(3):328–343.
- Hecq, A. and Velasquez-Gaviria, D. (2025). Non-causal and non-invertible ARMA models: Identification, estimation and application in equity portfolios. Journal of Time Series Analysis, 46(2):325–352.
- Hult, H. and Samorodnitsky, G. (2008). Tail probabilities for infinite series of regularly varying random vectors. Bernoulli, 14(3):838–864.
- Ibragimov, I. A. (1962). Some limit theorems for stationary processes. Theory of Probability and Its Applications, 7(4):349–382.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. The Annals of Statistics, 17(3):1217–1241.
- Lanne, M. and Saikkonen, P. (2011). Noncausal autoregressions for economic time series. Journal of Time Series Econometrics, 3(3).
- Lanne, M. and Saikkonen, P. (2013). Noncausal vector autoregression. Econometric Theory, 29(3):447–481.
- Liebscher, E. (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. Journal of Time Series Analysis, 26(5):669–689.
- Meitz, M. and Saikkonen, P. (2008). Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. Econometric Theory, 24(5):1291–1320.
- Meyn, S. and Tweedie, R. L. (2009). Markov Chains and Stochastic Stability. Cambridge University Press, Cambridge, 2 edition.

- Mikosch, T. and Wintenberger, O. (2024). Extreme Value Theory for Time Series: Models with Power-Law Tails. Springer Series in Operations Research and Financial Engineering. Springer, Cham.
- Mokkadem, A. (1988). Mixing properties of ARMA processes. Stochastic Processes and their Applications, 29(2):309–315.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3):703–708.
- Pham, T. D. and Tran, L. T. (1985). Some mixing properties of time series models. Stochastic Processes and their Applications, 19(2):297–303.
- Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. The Annals of Statistics, 22(4):2031–2050.
- Rio, E. (2017). Asymptotic Theory of Weakly Dependent Random Processes, volume 80 of Probability Theory and Stochastic Modelling. Springer, Berlin.
- Rosenblatt, M. (2000). Gaussian and Non-Gaussian Linear Time Series and Random Fields. Springer Series in Statistics. Springer, New York.
- Samorodnitsky, G. and Taqqu, M. S. (1994). Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Chapman & Hall, New York.
- Volkonskii, V. A. and Rozanov, Y. A. (1959). Some limit theorems for random functions. I. Theory of Probability and its Applications, 4(2):178–197.
- Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. The Annals of Probability, 22(1):94–116.

A Proofs

This appendix proves the results of the main text in the order in which they appear.

A.1 Proofs for Section 2

Proof of Lemma 1. The proof has two parts. We first show that already the first s directions $J_2^{-1}a_2, \dots, J_2^{-s}a_2$ span $\mathbb{R}^s$. We then split $Z_{2,t} = \Xi + \Xi'$, with Ξ an invertible linear image of the first s shocks $\eta_{t+1}, \dots, \eta_{t+s}$ and Ξ' independent of Ξ , so that $f_{Z_2} = f_\Xi * \mathbb{P}_{\Xi'}$ inherits positivity and continuity from f_η . We start with the spanning property. From

$$A^{-1}\Theta = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix} A^{-1},$$

the last s rows give

$$A_2\Theta = J_2A_2. \quad (\text{A.1})$$

Multiplying both sides by Θ on the right k times gives $A_2\Theta^k = J_2^kA_2$. Finally, multiplying both sides by $\mathbf{e}_1$ gives $A_2\Theta^k\mathbf{e}_1 = J_2^ka_2$ for $k \geq 0$. The companion matrix acts as a shift on the canonical basis, $\Theta\mathbf{e}_j = \mathbf{e}_{j+1} + \theta_j\mathbf{e}_1$ for $j = 1, \dots, p-1$. Starting from $\mathbf{e}_1$ and iterating, $\Theta^k\mathbf{e}_1 = \mathbf{e}_{k+1} + \sum_{j=1}^k \gamma_{k,j}\mathbf{e}_j$ for $0 \leq k \leq p-1$, for some real coefficients $\gamma_{k,j}$, the sum being empty for $k=0$. The $(k+1)$ -th column of $[\mathbf{e}_1, \Theta\mathbf{e}_1, \dots, \Theta^{p-1}\mathbf{e}_1]$ thus has a 1 in row $k+1$ and zeros in every row below it, so this matrix is upper triangular with unit diagonal. Hence it is invertible, and $\text{span}\{\Theta^k\mathbf{e}_1 : 0 \leq k \leq p-1\} = \mathbb{R}^p$. We now transport this basis through A_2 . As A^{-1} is invertible, A_2 has rank s and maps $\mathbb{R}^p$ onto $\mathbb{R}^s$, so applying A_2 to the above span and using $J_2^ka_2 = A_2\Theta^k\mathbf{e}_1$ gives

$$\text{span}\{a_2, J_2a_2, \dots, J_2^{p-1}a_2\} = A_2 \text{span}\{\mathbf{e}_1, \Theta\mathbf{e}_1, \dots, \Theta^{p-1}\mathbf{e}_1\} = A_2\mathbb{R}^p = \mathbb{R}^s.$$

By the Cayley–Hamilton theorem, every power J_2^k with $k \geq s$ is a linear combination of the lower powers $I, J_2, \dots, J_2^{s-1}$. The first s vectors therefore already suffice, that is, $\text{span}\{a_2, J_2a_2, \dots, J_2^{s-1}a_2\} = \mathbb{R}^s$. The series (2.6) involves the negative powers $J_2^{-1}a_2, \dots, J_2^{-s}a_2$ instead. Since J_2 is invertible, applying J_2^{-s} to this family gives

$$\text{span}\{J_2^{-1}a_2, \dots, J_2^{-s}a_2\} = J_2^{-s} \text{span}\{a_2, J_2a_2, \dots, J_2^{s-1}a_2\} = J_2^{-s}\mathbb{R}^s = \mathbb{R}^s. \quad (\text{A.2})$$

The loadings of the shocks $\eta_{t+1}, \dots, \eta_{t+s}$ therefore span $\mathbb{R}^s$ on their own, and we construct the density from these shocks. Collect the first s shocks of the series in the vector $\xi = (\eta_{t+1}, \dots, \eta_{t+s})' \in \mathbb{R}^s$. Its entries are independent with common density f_η , so ξ has density $f_\xi(x) = \prod_{j=1}^s f_\eta(x_j)$ for $x \in \mathbb{R}^s$. Split (2.6) accordingly as $Z_{2,t} = \Xi + \Xi'$, where

$$\Xi = -\sum_{j=0}^{s-1} J_2^{-(j+1)}a_2\eta_{t+j+1} = D_0\xi, \quad D_0 = -[J_2^{-1}a_2, \dots, J_2^{-s}a_2],$$

and $\Xi' = -\sum_{j \geq s} J_2^{-(j+1)} a_2 \eta_{t+j+1}$ collects the remaining shocks, so that Ξ' is independent of Ξ . The columns of $D_0 \in \mathbb{R}^{s \times s}$ span $\mathbb{R}^s$ by (A.2), so D_0 is invertible and $\Xi = D_0 \xi$ is a linear bijection of ξ . Hence Ξ has density

$$f_{\Xi}(w) = \frac{f_{\xi}(D_0^{-1}w)}{|\det D_0|} = \frac{1}{|\det D_0|} \prod_{j=1}^s f_{\eta}((D_0^{-1}w)_j), \quad w \in \mathbb{R}^s,$$

which is positive, continuous and bounded on $\mathbb{R}^s$ because f_{η} is positive, continuous and bounded on $\mathbb{R}$ by Assumption 2. Since Ξ and Ξ' are independent and Ξ admits a density, the law of $Z_{2,t} = \Xi + \Xi'$ is the convolution of the two laws, so $Z_{2,t}$ admits the density $f_{Z_2}(z) = \int_{\mathbb{R}^s} f_{\Xi}(z-r) \mathbb{P}_{\Xi'}(dr) = \mathbb{E}[f_{\Xi}(z - \Xi')]$, $z \in \mathbb{R}^s$, where $\mathbb{P}_{\Xi'}$ denotes the law of Ξ' . This is positive since $f_{\Xi} > 0$ everywhere. It is continuous by dominated convergence, f_{Ξ} being continuous and bounded, and it is bounded by $\sup_w f_{\Xi}(w) \leq (\sup f_{\eta})^s / |\det D_0|$. Finally, two continuous densities of the same law coincide Leb-almost everywhere, hence everywhere, so the continuous version is unique. $\square$

A.2 Proofs for Section 3.1

The three properties of Proposition 1 rest on the transition densities of $(\mathbf{X}_t)$. Start from $\mathbf{X}_0 = \mathbf{x} = (x_0, x_{-1}, \dots, x_{-p+1})'$. One iteration renews only the first coordinate of the state, $\mathbf{X}_1 = (X_1, x_0, x_{-1}, \dots, x_{-p+2})'$, which still contains $p-1$ coordinates of $\mathbf{x}$. More generally, for $k < p$, $\mathbf{X}_k = (X_k, \dots, X_1, x_0, \dots, x_{k-p+1})'$ shares $p-k$ coordinates with $\mathbf{x}$, so $\mathbf{X}_k$ is confined to an affine subspace of $\mathbb{R}^p$ fixed by the initial condition, and a set avoiding that subspace cannot be reached. Only after p iterations is the state $\mathbf{X}_p = (X_p, \dots, X_1)'$ made of entirely new observations. We therefore work with the n -step transitions for $n \geq p$.

To describe them, recall the notation $(x, \mathbf{y}_{-p})$. For $\mathbf{x} \in \mathbb{R}^p$ and $w = (w_1, \dots, w_n) \in \mathbb{R}^n$, define the successive states $\mathbf{s}_0 = \mathbf{x}$ and $\mathbf{s}_k = (w_k, (\mathbf{s}_{k-1})_{-p})$, $k = 1, \dots, n$. When $n \geq p$, $\mathbf{s}_n = (w_n, \dots, w_{n-p+1})'$ only depends on the last p coordinates of w , and, given $\mathbf{y} \in \mathbb{R}^p$, we set $(w_{n-p+1}, \dots, w_n) = (y_p, \dots, y_1)$, so that $\mathbf{s}_n = \mathbf{y}$.

Lemma A.1. *Under Assumptions 1 to 3, for every $n \geq p$ and $\mathbf{x} \in \mathbb{R}^p$, the n -step transition $P^n(\mathbf{x}, \cdot)$ admits the density, with respect to Leb,*

$$f_n(\mathbf{y} \mid \mathbf{x}) = |\det J_2|^n \frac{f_{Z_2}(A_2 \mathbf{y})}{f_{Z_2}(A_2 \mathbf{x})} \int_{\mathbb{R}^{n-p}} \prod_{k=1}^n f_{\eta}(w_k - \boldsymbol{\theta}' \mathbf{s}_{k-1}) dw_1 \cdots dw_{n-p}, \quad \mathbf{y} \in \mathbb{R}^p, \quad (\text{A.3})$$

with no integration when $n = p$. The function $(\mathbf{x}, \mathbf{y}) \mapsto f_n(\mathbf{y} \mid \mathbf{x})$ is finite, positive and jointly continuous on $\mathbb{R}^p \times \mathbb{R}^p$.

Proof of Lemma A.1. By construction of $\mathbb{P}_{\mathbf{x}}$ from the kernel P , $\mathbf{X}_k = \mathbf{s}_k$ with $w_k = X_k$, and, for every $B \in \mathcal{B}(\mathbb{R}^p)$,

$$P^n(\mathbf{x}, B) = \int_{\mathbb{R}^n} \mathbf{1}_B(\mathbf{s}_n) \prod_{k=1}^n f(w_k \mid \mathbf{s}_{k-1}) dw_1 \cdots dw_n.$$

Since $\mathbf{s}_k = (w_k, (\mathbf{s}_{k-1})_{-p})$, (2.7) gives

$$f(w_k \mid \mathbf{s}_{k-1}) = |\det J_2| \frac{f_{Z_2}(A_2 \mathbf{s}_k) f_{\eta}(w_k - \boldsymbol{\theta}' \mathbf{s}_{k-1})}{f_{Z_2}(A_2 \mathbf{s}_{k-1})},$$

and the ratios of f_{Z_2} telescope, so that

$$\prod_{k=1}^n f(w_k \mid \mathbf{s}_{k-1}) = |\det J_2|^n \frac{f_{Z_2}(A_2 \mathbf{s}_n)}{f_{Z_2}(A_2 \mathbf{x})} \prod_{k=1}^n f_{\eta}(w_k - \boldsymbol{\theta}' \mathbf{s}_{k-1}).$$

As $n \geq p$, $\mathbf{s}_n$ is obtained from $(w_{n-p+1}, \dots, w_n)$ by a permutation of coordinates. Integrating first over $w_1, \dots, w_{n-p}$, by Tonelli's theorem, gives $P^n(\mathbf{x}, B) = \int_B f_n(\mathbf{y} \mid \mathbf{x}) d\text{Leb}(\mathbf{y})$ with f_n as in (A.3).

We now prove the regularity of f_n . When $n = p$, (A.3) is a finite product of positive continuous functions of $(\mathbf{x}, \mathbf{y})$, since each $\mathbf{s}_k$ is linear in $(\mathbf{x}, \mathbf{y})$ and f_{Z_2}, f_{η} are positive and continuous (Lemma 1 and Assumption 2). When $n > p$, change variables in the integral through $u_k = w_k - \boldsymbol{\theta}' \mathbf{s}_{k-1}$, $k = 1, \dots, n - p$. Since $\mathbf{s}_{k-1}$ only depends on $\mathbf{x}$ and $w_1, \dots, w_{k-1}$, this is a triangular affine bijection of $\mathbb{R}^{n-p}$ with unit Jacobian, whose inverse is continuous in $(u, \mathbf{x})$. The integrand becomes

$$\prod_{k=1}^{n-p} f_{\eta}(u_k) \prod_{k=n-p+1}^n f_{\eta}(w_k - \boldsymbol{\theta}' \mathbf{s}_{k-1}),$$

where, for $k > n - p$, w_k is a coordinate of $\mathbf{y}$ and $\mathbf{s}_{k-1}$ is a continuous function of $(u, \mathbf{x}, \mathbf{y})$. This integrand is positive, continuous in $(\mathbf{x}, \mathbf{y})$ for every u , and bounded by $(\sup f_{\eta})^p \prod_{k \leq n-p} f_{\eta}(u_k)$, which is integrable and does not depend on $(\mathbf{x}, \mathbf{y})$. By dominated convergence, the integral in (A.3) is finite, positive and jointly continuous, and so is f_n . $\square$

Proof of Proposition 1. (ι) Let $B \in \mathcal{B}(\mathbb{R}^p)$ with $\text{Leb}(B) > 0$, and let $\mathbf{x} \in \mathbb{R}^p$. By Lemma A.1, the density $f_p(\cdot \mid \mathbf{x})$ is positive, so $\mathbb{P}_{\mathbf{x}}(\mathbf{X}_p \in B) = \int_B f_p(\mathbf{y} \mid \mathbf{x}) d\text{Leb}(\mathbf{y}) > 0$. Finally, with $\tau_B = \inf\{n \geq 1 : \mathbf{X}_n \in B\}$ the first hitting time of B , $\{\mathbf{X}_p \in B\} \subseteq \{\tau_B \leq p\}$, whence $\mathbb{P}_{\mathbf{x}}(\tau_B < \infty) \geq \mathbb{P}_{\mathbf{x}}(\mathbf{X}_p \in B) > 0$, which is Leb-irreducibility.

(ι) Let $C \in \mathcal{B}(\mathbb{R}^p)$ be nonempty and bounded and let $\mathcal{K} \subset \mathbb{R}^p$ be any compact set with $\text{Leb}(\mathcal{K}) > 0$. The closure $\overline{C}$ is compact, hence so is $\overline{C} \times \mathcal{K}$. The p -step density is strictly positive and jointly continuous (Lemma A.1), so it attains its infimum on that compact set and $\delta := \inf_{(\mathbf{x}, \mathbf{y}) \in \overline{C} \times \mathcal{K}} f_p(\mathbf{y} \mid \mathbf{x}) > 0$.

Let $\mathbf{x} \in C$ and $B \in \mathcal{B}(\mathbb{R}^p)$. Restricting the integral to $B \cap \mathcal{K}$ and bounding the integrand below by δ ,

$$P^p(\mathbf{x}, B) \geq \int_{B \cap \mathcal{K}} f_p(\mathbf{y} \mid \mathbf{x}) d\text{Leb}(\mathbf{y}) \geq \delta \text{Leb}(B \cap \mathcal{K}) =: \nu_p(B).$$

The set function ν_p is a measure on $\mathcal{B}(\mathbb{R}^p)$, and it is non-trivial since $\nu_p(\mathcal{K}) = \delta \text{Leb}(\mathcal{K}) > 0$. Hence C is ν_p -small in the sense of [Meyn and Tweedie \(2009\)](#).

($\mu\mu$) Let C be a closed ball of $\mathbb{R}^p$ with positive radius. Fix $n \geq p$. The n -step density is strictly positive and jointly continuous ([Lemma A.1](#)), so it attains a strictly positive infimum on the compact set $C \times C$, $\delta_n := \inf_{(\mathbf{x}, \mathbf{y}) \in C \times C} f_n(\mathbf{y} \mid \mathbf{x}) > 0$. Consequently, for every $\mathbf{x} \in C$ and $B \in \mathcal{B}(\mathbb{R}^p)$,

$$P^n(\mathbf{x}, B) \geq \int_{B \cap C} f_n(\mathbf{y} \mid \mathbf{x}) d\text{Leb}(\mathbf{y}) \geq \delta_n \text{Leb}(B \cap C) = \delta_n \nu(B), \quad \nu(B) := \text{Leb}(B \cap C).$$

The measure ν satisfies $\nu(C) = \text{Leb}(C) > 0$ and, unlike δ_n , does not depend on n . Taking $M = p$ and $\nu_M = \delta_p \nu$, C is ν_n -small with $\nu_n = (\delta_n / \delta_p) \nu_M$ for every $n \geq p$. The set over which the greatest common divisor, d , is taken therefore contains $\{p, p+1\}$, so it divides both p and $p+1$, hence divides $(p+1) - p = 1$. Thus $d = 1$ and $(\mathbf{X}_t)$ is aperiodic. $\square$

A.3 Technical results for [Section 4](#)

This subsection states and proves the technical lemmas used in [Section 4](#) and in [Section A.4](#).

A.3.1 Decay and nondegeneracy of the loadings

Two properties of the loadings are used below. They decay geometrically at a rate we can name, and any s consecutive ones span $\mathbb{R}^s$. The spectral radius of J_2^{-1} is λ_*^{-1} , so Gelfand's formula gives $\|J_2^{-j}\|^{1/j} \rightarrow \lambda_*^{-1}$ and, for every $\lambda \in (1, \lambda_*)$,

$$C_\lambda = \sup_{j \geq 0} \lambda^j \|J_2^{-j}\| < \infty, \tag{A.4}$$

since $\lambda \|J_2^{-j}\|^{1/j} \rightarrow \lambda / \lambda_* < 1$. The loadings then satisfy

$$d_1 = -J_2^{-1} a_2, \quad d_{j+1} = J_2^{-1} d_j, \quad \|d_j\| \leq \|a_2\| C_\lambda \lambda^{-j} \quad (j \geq 1), \tag{A.5}$$

and, for $\lambda_0 \in [1, \lambda_*)$, choosing $\lambda \in (\lambda_0, \lambda_*)$ gives $\lambda_0^j \|d_j\| \leq \|a_2\| C_\lambda (\lambda_0 / \lambda)^j$, whence

$$\sup_{j \geq 1} \lambda_0^j \|d_j\| < \infty \quad \text{and} \quad \sum_{j \geq 1} (\lambda_0^j \|d_j\|)^\kappa < \infty \quad \text{for every } \kappa > 0. \tag{A.6}$$

Consider the matrices $D_n = [d_{n+1}, \dots, d_{n+s}] = J_2^{-n} D_0 \in \mathbb{R}^{s \times s}$, $n \geq 0$. By (A.2) the columns of $D_0 = -[J_2^{-1} a_2, \dots, J_2^{-s} a_2]$ span $\mathbb{R}^s$, so D_0 is invertible and hence so is every D_n .

A.3.2 Tail behavior of the error density

Let $\ell_\alpha(x) = x^{-(\alpha+1)} \ell(x)$ for $x > 0$, a regularly varying function of index $-(\alpha+1)$. By the uniform convergence theorem for regularly varying functions (Mikosch and Wintenberger, 2024),

$$\frac{\ell_\alpha(\mu x)}{\ell_\alpha(x)} \longrightarrow \mu^{-(\alpha+1)}, \quad x \rightarrow \infty, \quad (\text{A.7})$$

uniformly in μ on compact subsets of $(0, \infty)$. Since $f_\eta(x) = |\psi_s| f_\varepsilon(-\psi_s x)$, Assumption 3 and (A.7) give

$$f_\eta(x) \sim g_\pm \ell_\alpha(|x|), \quad x \rightarrow \pm\infty, \quad g_\pm = |\psi_s|^{-\alpha} c_{\mp \text{sgn}(\psi_s)} > 0. \quad (\text{A.8})$$

The following bound states that f_η cannot change by an unbounded factor over a bounded distance.

Lemma A.2. *Under Assumptions 2 and 3, for every $\delta \geq 0$,*

$$H_\eta(\delta) = \sup_{|u-v| \leq \delta} \frac{f_\eta(u)}{f_\eta(v)} < \infty. \quad (\text{A.9})$$

The function H_η is nondecreasing and $H_\eta \geq 1$.

Proof of Lemma A.2. The supremum in (A.9) ranges over all of $\mathbb{R}$, while Assumption 3 describes f_η only near $\pm\infty$. We therefore treat large and bounded arguments separately. Fix $\delta > 0$ and recall from (A.8) that $f_\eta(x)/\ell_\alpha(|x|) \rightarrow g_\pm > 0$ as $x \rightarrow \pm\infty$. We first show that the ratio tends to one for large arguments. Let $x \geq 2\delta$ and $y \in \mathbb{R}$ with $|y - x| \leq \delta$. Then $y \geq x - \delta \geq \delta > 0$, and $\mu = y/x$ lies in $[1 - \delta/x, 1 + \delta/x] \subseteq [1/2, 3/2]$. Write

$$\frac{f_\eta(y)}{f_\eta(x)} = \frac{f_\eta(y)}{\ell_\alpha(y)} \cdot \frac{\ell_\alpha(\mu x)}{\ell_\alpha(x)} \cdot \frac{\ell_\alpha(x)}{f_\eta(x)}.$$

As $x \rightarrow \infty$ the first factor tends to g_+ uniformly in $|y - x| \leq \delta$, because $y \geq x - \delta$, and the third tends to $1/g_+$. By (A.7), $\ell_\alpha(\mu x)/\ell_\alpha(x) - \mu^{-(\alpha+1)} \rightarrow 0$ uniformly in $\mu \in [1/2, 3/2]$, and $\mu^{-(\alpha+1)} \rightarrow 1$ uniformly in $\mu \in [1 - \delta/x, 1 + \delta/x]$. Hence $\sup_{|y-x| \leq \delta} f_\eta(y)/f_\eta(x) \rightarrow 1$ as $x \rightarrow +\infty$. The same argument applies as $x \rightarrow -\infty$, with $|x|$, $|y|$ and g_- , since then $y \leq x + \delta < 0$. There is therefore $u_0 \geq 2\delta$ such that $f_\eta(y) \leq 2f_\eta(x)$ whenever $|x| \geq u_0$ and $|y - x| \leq \delta$. It remains to treat bounded arguments. If $|x| < u_0$ and $|y - x| \leq \delta$, then x and y both lie in the compact interval $\mathcal{K} = [-u_0 - \delta, u_0 + \delta]$. By Assumption 2 f_η is continuous and positive, so $\min_{\mathcal{K}} f_\eta > 0$ and $f_\eta(y)/f_\eta(x) \leq \sup f_\eta / \min_{\mathcal{K}} f_\eta$. Together with the bound for large arguments, this gives $H_\eta(\delta) \leq \max(2, \sup f_\eta / \min_{\mathcal{K}} f_\eta) < \infty$. $\square$

A.3.3 Densities of partial sums of the noncausal state

For an index set $\mathcal{S} \subseteq \{1, 2, \dots\}$, write $\eta_{\mathcal{S}} = (\eta_j)_{j \in \mathcal{S}}$ for the corresponding family of future shocks, $Y_{\mathcal{S}} = \sum_{j \in \mathcal{S}} d_j \eta_j$, and $\mathcal{I}_n = \{n+1, \dots, n+s\}$ for $n \geq 0$, so that $Y_{\mathcal{S}}$ is the partial sum of the series of $Z_{2,0}$ over $\mathcal{S}$ and is a function of $\eta_{\mathcal{S}}$ only, and $\eta_{\mathcal{I}_n} = (\eta_{n+1}, \dots, \eta_{n+s})'$ is a block of s consecutive shocks. By [Assumption 3](#), $\mathbb{E}|\eta_t|^\gamma < \infty$ for every $\gamma \in (0, \alpha)$, so [\(A.6\)](#) and [Brockwell and Davis \(1991, Proposition 13.3.1\)](#) show that the series of $Z_{2,0}$, and hence $Y_{\mathcal{S}}$, converges absolutely almost surely. Requiring $\mathcal{I}_n \subseteq \mathcal{S}$ is what makes $Y_{\mathcal{S}}$ nondegenerate, since D_n is invertible.

Lemma A.3. *Let $n \geq 0$ and $\mathcal{S} \subseteq \{1, 2, \dots\}$ with $\mathcal{I}_n \subseteq \mathcal{S}$. Then $Y_{\mathcal{S}}$ has a density $f_{\mathcal{S}}$ which is positive, continuous and bounded by $(\sup f_{\eta})^s / |\det D_n|$, and which satisfies*

$$f_{\mathcal{S}}(w) \leq H_{\eta}(\|D_n^{-1}\| \|y - w\|)^s f_{\mathcal{S}}(y), \quad y, w \in \mathbb{R}^s.$$

Proof of Lemma A.3. Split $\mathcal{S} = \mathcal{I}_n \cup (\mathcal{S} \setminus \mathcal{I}_n)$ and let $\xi = \eta_{\mathcal{I}_n}$ and $Y_{\mathcal{T}}$, with $\mathcal{T} = \mathcal{S} \setminus \mathcal{I}_n$. They are functions of shocks with disjoint index sets, hence independent, and $Y_{\mathcal{S}} = D_n \xi + Y_{\mathcal{T}}$ since the columns of D_n are $d_{n+1}, \dots, d_{n+s}$. As in the proof of [Lemma 1](#), with D_n in place of D_0 , the block $D_n \xi$ has density $f_{D_n \xi}(w) = |\det D_n|^{-1} \prod_{i=1}^s f_{\eta}((D_n^{-1} w)_i)$, and $f_{\mathcal{S}}(y) = \mathbb{E}[f_{D_n \xi}(y - Y_{\mathcal{T}})]$ is a positive and continuous density of $Y_{\mathcal{S}}$, bounded by $(\sup f_{\eta})^s / |\det D_n|$.

It remains to prove the ratio bound. Let $y, w, y_0 \in \mathbb{R}^s$ and set $\delta = \|D_n^{-1}\| \|y - w\|$. For each $i = 1, \dots, s$, linearity and the definition of the operator norm give

$$|(D_n^{-1}(w - y_0))_i - (D_n^{-1}(y - y_0))_i| = |(D_n^{-1}(w - y))_i| \leq \|D_n^{-1}(w - y)\| \leq \delta,$$

so $f_{\eta}((D_n^{-1}(w - y_0))_i) \leq H_{\eta}(\delta) f_{\eta}((D_n^{-1}(y - y_0))_i)$ by [\(A.9\)](#). Multiplying these s inequalities and using the product form of $f_{D_n \xi}$, $f_{D_n \xi}(w - y_0) \leq H_{\eta}(\delta)^s f_{D_n \xi}(y - y_0)$. Taking $y_0 = Y_{\mathcal{T}}$ and expectations gives the claimed bound. $\square$

We apply it to $R_k = \sum_{j \geq 1, j \neq k} d_j \eta_j = Z_{2,0} - d_k \eta_k$ of [Section 4.1](#), with density f_{R_k} . Its index set contains $\mathcal{I}_0$ if $k > s$ and $\mathcal{I}_s$ if $k \leq s$, so only D_0 and D_s occur and the bounds of [Lemma A.3](#) can be made uniform in k . Let C_R be a bound valid for every f_{R_k} and $H(\delta)$ a ratio bound valid for every f_{R_k} at distance δ ; we may take

$$C_R = \max_{n \in \{0, s\}} \frac{(\sup f_{\eta})^s}{|\det D_n|}, \quad H(\delta) = \max_{n \in \{0, s\}} H_{\eta}(\|D_n^{-1}\| \delta)^s, \quad \delta \geq 0.$$

The function H is nondecreasing and $H \geq 1$. Uniformity in k matters, since W sums over all horizons and a bound deteriorating with k would be of no use.

Lemma A.4. *For all $k \geq 1$ and $y, w, z \in \mathbb{R}^s$,*

$$f_{R_k} \leq C_R, \quad f_{R_k}(w) \leq H(\|y - w\|) f_{R_k}(y), \quad (\text{A.10})$$

and

$$f_{Z_2}(z) = \int_{\mathbb{R}} f_{\eta}(u) f_{R_k}(z - d_k u) du, \quad (\text{A.11})$$

$$f_{R_{k+1}}(y) = |\det J_2| \int_{\mathbb{R}} f_{\eta}(u) f_{R_k}(J_2 y + a_2 u) du, \quad (\text{A.12})$$

$$f_{R_1}(y) = |\det J_2| f_{Z_2}(J_2 y). \quad (\text{A.13})$$

Proof of Lemma A.4. Lemma A.3 applies to R_k with $n = 0$ if $k > s$ and with $n = s$ if $k \leq s$, which gives (A.10) by definition of C_R and H . We now prove the convolution identity (A.11). We have $Z_{2,0} = d_k \eta_k + R_k$ with R_k independent of η_k , so for measurable $h \geq 0$, the translation $z = y + d_k u$ gives

$$\mathbb{E}h(Z_{2,0}) = \int_{\mathbb{R}} f_{\eta}(u) \int_{\mathbb{R}^s} h(y + d_k u) f_{R_k}(y) dy du = \int_{\mathbb{R}^s} h(z) \int_{\mathbb{R}} f_{\eta}(u) f_{R_k}(z - d_k u) du dz.$$

For the recursion (A.12), we split off the first shock and reindex the others with $j = i + 1$. By (A.5),

$$R_{k+1} = d_1 \eta_1 + \sum_{i \geq 1, i \neq k} d_{i+1} \eta_{i+1} = J_2^{-1}(S_k - a_2 \eta_1), \quad S_k = \sum_{i \geq 1, i \neq k} d_i \eta_{i+1}.$$

The variable S_k is the same function of $(\eta_2, \eta_3, \dots)$ as R_k is of $(\eta_1, \eta_2, \dots)$, so it has density f_{R_k} and is independent of η_1 . For measurable $h \geq 0$, the substitution $r = J_2 y + a_2 u$ gives

$$\begin{aligned} \mathbb{E}h(R_{k+1}) &= \int_{\mathbb{R}} f_{\eta}(u) \int_{\mathbb{R}^s} h(J_2^{-1}(r - a_2 u)) f_{R_k}(r) dr du \\ &= \int_{\mathbb{R}^s} h(y) |\det J_2| \int_{\mathbb{R}} f_{\eta}(u) f_{R_k}(J_2 y + a_2 u) du dy. \end{aligned}$$

Finally, for (A.13), (A.5) gives $d_{i+1} = J_2^{-1} d_i$, so $R_1 = J_2^{-1} \sum_{i \geq 1} d_i \eta_{i+1}$, and the last sum has the law of $Z_{2,0}$. Hence R_1 has density $|\det J_2| f_{Z_2}(J_2 y)$. $\square$

A.3.4 The conditional law of the future shocks

We collect here the properties of the conditional laws of the future shocks. By (A.11), $p_k(\cdot | z)$ is a probability density for every z , and since $(u, z) \mapsto p_k(u | z)$ is continuous, $z \mapsto \mathbb{E}[h(\eta_k) | Z_{2,0} = z]$ is measurable. Since $Z_{2,0} = d_k \eta_k + R_k$ with R_k independent of η_k , $(\eta_k, Z_{2,0})$ has joint density $f_{\eta}(u) f_{R_k}(z - d_k u)$, so that $p_k(\cdot | z)$ is indeed the conditional density of η_k given $Z_{2,0} = z$, which gives (4.3).

The following properties are used repeatedly. If $\mathbb{E}[h(\eta_t)] < \infty$, the bound $f_{R_k} \leq C_R$ gives

$$\mathbb{E}[h(\eta_k) | Z_{2,0} = z] \leq \frac{C_R \mathbb{E}[h(\eta_t)]}{f_{Z_2}(z)}, \quad z \in \mathbb{R}^s, \quad (\text{A.14})$$

and $z \mapsto \mathbb{E}[h(\eta_k) \mid Z_{2,0} = z]$ is continuous, since f_{Z_2} is continuous and positive and the numerator, $z \mapsto \int_{\mathbb{R}} h(u) f_{\eta}(u) f_{R_k}(z - d_k u) du$, is continuous by dominated convergence with dominating function $C_R h f_{\eta}$. Finally, for $k = 1$, (A.13) and $J_2 d_1 = -a_2$ give $f_{R_1}(z - d_1 u) = |\det J_2| f_{Z_2}(J_2 z + a_2 u)$, which yields (4.4).

Lemma A.5. *Let $\lambda_0 \in (1, \lambda_*)$ and $M > 0$. There exists $C_M < \infty$ such that, for every $k \geq 1$, $z \in \mathbb{R}^s$ and Borel subset I of the interval $[-(1+M)\lambda_0^k, (1+M)\lambda_0^k]$,*

$$\mathbb{P}(\eta_k \in I \mid Z_{2,0} = z) \leq C_M \mathbb{P}(\eta_t \in I).$$

Proof of Lemma A.5. The density $p_k(\cdot \mid z)$ carries the factor $1/f_{Z_2}(z)$, which diverges as $\|z\| \rightarrow \infty$, so a bound uniform in z must come from a matching lower bound on $f_{Z_2}(z)$. Let δ be large enough that, for every k , two values u, v of the k -th shock in the window give arguments $z - d_k u$ and $z - d_k v$ of f_{R_k} at distance at most δ ; by (A.6) we may take $\delta = 2(1+M) \sup_{j \geq 1} \lambda_0^j \|d_j\| < \infty$ and $C_M = H(\delta)/\mathbb{P}(|\eta_t| \leq 1)$, where $\mathbb{P}(|\eta_t| \leq 1) > 0$ since f_{η} is positive. Neither depends on k or on z , as the statement requires. Fix $k \geq 1$ and $z \in \mathbb{R}^s$. We first compare the density of the remaining shocks at two points of the window. For $|u|, |v| \leq (1+M)\lambda_0^k$, $\|(z - d_k u) - (z - d_k v)\| = \|d_k\| \|u - v\| \leq 2(1+M)\lambda_0^k \|d_k\| \leq \delta$, so $f_{R_k}(z - d_k u) \leq H(\delta) f_{R_k}(z - d_k v)$ by (A.10), H being nondecreasing. We then deduce a lower bound on $f_{Z_2}(z)$. Fix u with $|u| \leq (1+M)\lambda_0^k$. Restricting the integral in (A.11) to $|v| \leq (1+M)\lambda_0^k$ and using the previous comparison,

$$\begin{aligned} f_{Z_2}(z) &\geq \int_{|v| \leq (1+M)\lambda_0^k} f_{\eta}(v) f_{R_k}(z - d_k v) dv \\ &\geq \frac{f_{R_k}(z - d_k u)}{H(\delta)} \mathbb{P}(|\eta_t| \leq (1+M)\lambda_0^k) \geq \frac{f_{R_k}(z - d_k u)}{C_M}, \end{aligned}$$

the last inequality using $(1+M)\lambda_0^k \geq 1$. Dividing by $f_{Z_2}(z)$ gives $p_k(u \mid z) \leq C_M f_{\eta}(u)$ for every u with $|u| \leq (1+M)\lambda_0^k$, and integrating over I proves the lemma. $\square$

Lemma A.6. *Let $\kappa \in (0, 1]$ with $\kappa < \alpha$. For every $z \in \mathbb{R}^s$,*

$$\|z\|^\kappa \leq \sum_{k \geq 1} \|d_k\|^\kappa \mathbb{E}[|\eta_k|^\kappa \mid Z_{2,0} = z].$$

Proof of Lemma A.6. We first show that the right-hand side is a continuous function of z . Let $Q(z) = \sum_{k \geq 1} \|d_k\|^\kappa \mathbb{E}[|\eta_k|^\kappa \mid Z_{2,0} = z]$. Since $\mathbb{E}|\eta_t|^\kappa < \infty$ by Assumption 3, (A.14) with $h(u) = |u|^\kappa$ shows that each term is continuous and bounded by $\|d_k\|^\kappa C_R \mathbb{E}|\eta_t|^\kappa / f_{Z_2}(z)$. On a compact set f_{Z_2} is bounded below by a positive constant and $\sum_k \|d_k\|^\kappa < \infty$ by (A.6), so the series converges uniformly on compact sets and Q is finite and continuous.

We then establish the inequality in integrated form. As the series of $Z_{2,0}$ converges absolutely, $\|Z_{2,0}\| \leq \sum_k \|d_k\| |\eta_k|$, so the subadditivity of $y \mapsto y^\kappa$ on $[0, \infty)$ gives $\|Z_{2,0}\|^\kappa \leq \sum_k \|d_k\|^\kappa |\eta_k|^\kappa$ almost surely. Let $B \subseteq \mathbb{R}^s$ be a bounded Borel set. Multiplying by $\mathbf{1}\{Z_{2,0} \in B\}$, taking expectations and using (4.3) with $h(u) = |u|^\kappa$,

$$\int_B \|z\|^\kappa f_{Z_2}(z) dz \leq \sum_{k \geq 1} \|d_k\|^\kappa \mathbb{E}[|\eta_k|^\kappa \mathbf{1}\{Z_{2,0} \in B\}] = \int_B Q(z) f_{Z_2}(z) dz,$$

both integrals being finite since B is bounded and Q continuous. Taking $B = \{z : \|z\| \leq n, \|z\|^\kappa > Q(z)\}$, the function $(\|z\|^\kappa - Q(z))f_{Z_2}(z)$ is positive on B and its integral over B is nonpositive, so B has Lebesgue measure zero. Letting $n \rightarrow \infty$, the set $O = \{z : \|z\|^\kappa > Q(z)\}$ has Lebesgue measure zero. It is open, both functions being continuous, and a nonempty open set has positive Lebesgue measure, so O is empty. $\square$

A.4 Proofs for Section 3.2 and Section 4

Proof of Lemma 2. Let $\rho_1 \in (\lambda_\phi^{-1}, 1)$, which can be taken arbitrarily close to λ_ϕ^{-1} , as Remark 2 requires. Since λ_ϕ^{-1} is the spectral radius of J_1 , Gelfand's formula gives $k_0 \geq 1$ with $\|J_1^{k_0}\| \leq \rho_1^{k_0}$ for the Euclidean operator norm. Set $\|v\|_* = \sum_{k=0}^{k_0-1} \rho_1^{-k} \|J_1^k v\|$, a norm since its $k=0$ term is $\|v\|$. Then

$$\|J_1 v\|_* = \rho_1 \left(\sum_{k=1}^{k_0-1} \rho_1^{-k} \|J_1^k v\| + \rho_1^{-k_0} \|J_1^{k_0} v\| \right) \leq \rho_1 \left(\sum_{k=1}^{k_0-1} \rho_1^{-k} \|J_1^k v\| + \|v\| \right) = \rho_1 \|v\|_*. \quad \square$$

Proof of Lemma 3. Let $z = A_2 \mathbf{x}$. By (2.5) and (4.4), $Z_{2,1} = J_2 z + a_2 \eta_1$ with η_1 of density $p_1(\cdot | z)$, so $\mathbb{E}_{\mathbf{x}}[m(Z_{2,1})] = \int_{\mathbb{R}} p_1(u | z) m(J_2 z + a_2 u) du$. Write $w = J_2 z + a_2 u$. By (4.4) and (4.1),

$$p_1(u | z) = |\det J_2| \frac{f_{Z_2}(w)}{f_{Z_2}(z)} f_\eta(u), \quad m(w) = \frac{1}{f_{Z_2}(w)} \int_{\mathbb{R}} h(v) f_\eta(v) f_{R_k}(w - d_k v) dv,$$

and $f_{Z_2}(w) > 0$ cancels in the product. Exchanging the order of integration gives

$$\mathbb{E}_{\mathbf{x}}[m(Z_{2,1})] = \frac{1}{f_{Z_2}(z)} \int_{\mathbb{R}} h(v) f_\eta(v) \left[|\det J_2| \int_{\mathbb{R}} f_\eta(u) f_{R_k}(J_2 z + a_2 u - d_k v) du \right] dv.$$

By (A.5), $d_k = J_2 d_{k+1}$, so the argument of f_{R_k} equals $J_2(z - d_{k+1}v) + a_2 u$, and (A.12) at $y = z - d_{k+1}v$ shows the bracket to be $f_{R_{k+1}}(z - d_{k+1}v)$. The right-hand side is therefore $\int_{\mathbb{R}} h(v) p_{k+1}(v | z) dv = \mathbb{E}[h(\eta_{k+1}) | Z_{2,0} = z]$. $\square$

Proof of Proposition 2. All three statements rest on the same computation, which bounds the k -th term of W by q^k times a fixed moment. We first bound the unconditional means. On $\{|u| > \lambda_0^k\}$ we have $|u|^\kappa = |u|^\gamma |u|^{-(\gamma-\kappa)} \leq |u|^\gamma \lambda_0^{-k(\gamma-\kappa)}$, and $h_k(u) \leq |u|^\kappa \mathbf{1}\{|u| > \lambda_0^k\}$,

so

$$\mathbb{E}[h_k(\eta_t)] \leq \lambda_0^{-k(\gamma-\kappa)} \mathbb{E}|\eta_t|^\gamma, \quad \mathbb{P}(|\eta_t| > \lambda_0^k) \leq \lambda_0^{-k\gamma} \mathbb{E}|\eta_t|^\gamma, \quad (\text{A.15})$$

where $\mathbb{E}|\eta_t|^\gamma < \infty$ by [Assumption 3](#). In particular $\omega^{-k} \mathbb{E}[h_k(\eta_t)] \leq q^k \mathbb{E}|\eta_t|^\gamma$, which is where the constraint $\omega > \lambda_0^{-(\alpha-\kappa)}$ of (3.3) is used. We now deduce finiteness and continuity. By (A.14) with $h = h_k$, each term of W is continuous and

$$\omega^{-k} \mathbb{E}[h_k(\eta_k) \mid Z_{2,0} = z] \leq \frac{C_R \omega^{-k} \mathbb{E}[h_k(\eta_t)]}{f_{Z_2}(z)} \leq \frac{C_R \mathbb{E}|\eta_t|^\gamma}{f_{Z_2}(z)} q^k.$$

Summing over k gives $W(z) \leq C_R \mathbb{E}|\eta_t|^\gamma q / \{(1-q)f_{Z_2}(z)\} < \infty$. On a compact set $\mathcal{K}$, f_{Z_2} is bounded below by a positive constant, so the display gives a summable bound uniform on $\mathcal{K}$; the series converges uniformly on $\mathcal{K}$ and W is continuous. For (4.5), we compare $|u|^\kappa$ with $h_k(u)$. For every $u \in \mathbb{R}$,

$$|u|^\kappa \leq h_k(u) + \lambda_0^{k\kappa}, \quad (\text{A.16})$$

which is clear if $|u| \leq \lambda_0^k$ and follows from subadditivity otherwise. Evaluating at $u = \eta_k$ and taking conditional expectations gives $\mathbb{E}[|\eta_k|^\kappa \mid Z_{2,0} = z] \leq \mathbb{E}[h_k(\eta_k) \mid Z_{2,0} = z] + \lambda_0^{k\kappa}$. Let $\bar{d} = \sup_{j \geq 1} \|d_j\|$ and $C_2 = \sum_{j \geq 1} (\lambda_0^j \|d_j\|)^\kappa$, both finite by (A.6), which is where the constraint $\lambda_0 < \lambda_*$ is used. By [Lemma A.6](#) and (A.16), and then $\|d_k\| \leq \bar{d}$ and $\omega^{-k} \geq 1$,

$$\|z\|^\kappa \leq \sum_{k \geq 1} \|d_k\|^\kappa \mathbb{E}[h_k(\eta_k) \mid Z_{2,0} = z] + C_2 \leq \bar{d}^\kappa W(z) + C_2,$$

which is (4.5) with $C_W = \max(\bar{d}^\kappa, C_2)$. Finally, (4.6) follows from (4.3) with $B = \mathbb{R}^s$ and (A.15), since

$$\begin{aligned} \int_{\mathbb{R}^s} W f_{Z_2} &= \sum_{k \geq 1} \omega^{-k} \int_{\mathbb{R}^s} \mathbb{E}[h_k(\eta_k) \mid Z_{2,0} = z] f_{Z_2}(z) dz \\ &= \sum_{k \geq 1} \omega^{-k} \mathbb{E}[h_k(\eta_t)] \leq \mathbb{E}|\eta_t|^\gamma \sum_{k \geq 1} q^k < \infty. \end{aligned} \quad \square$$

Proof of Proposition 3. Let $M = (\rho_2/\omega - 1)^{-1/\kappa}$, which is positive since $\rho_2 > \omega$ and satisfies $\omega(1 + M^{-\kappa}) = \rho_2$, and let C_M be the constant of [Lemma A.5](#) for this M . Fix $\mathbf{x} \in \mathbb{R}^p$ and let $z = A_2 \mathbf{x}$. We start with a pointwise bound. For $k \geq 1$ let $I_{k+1} = \{u \in \mathbb{R} : \lambda_0^k < |u| \leq (1 + M)\lambda_0^{k+1}\}$. We claim that, for every $u \in \mathbb{R}$,

$$h_k(u) \leq (1 + M^{-\kappa}) h_{k+1}(u) + \lambda_0^{\kappa(k+1)} \mathbf{1}\{u \in I_{k+1}\}. \quad (\text{A.17})$$

Write $\tau_1 = \lambda_0^k < \tau_2 = \lambda_0^{k+1}$, so that $h_k(u) = (|u| - \tau_1)_+^\kappa$ and $h_{k+1}(u) = (|u| - \tau_2)_+^\kappa$. If $|u| \leq \tau_1$ the left-hand side vanishes. If $\tau_1 < |u| \leq \tau_2$ then $u \in I_{k+1}$ and the left-hand side is at most $(\tau_2 - \tau_1)^\kappa \leq \tau_2^\kappa$. If $|u| > \tau_2$, subadditivity gives $(|u| - \tau_1)^\kappa \leq (|u| - \tau_2)^\kappa + (\tau_2 - \tau_1)^\kappa \leq h_{k+1}(u) + \tau_2^\kappa$, and either

$|u| \leq (1 + M)\tau_2$, in which case $u \in I_{k+1}$, or $|u| > (1 + M)\tau_2$, in which case $\tau_2^\kappa = M^{-\kappa}(M\tau_2)^\kappa < M^{-\kappa}(|u| - \tau_2)^\kappa = M^{-\kappa}h_{k+1}(u)$. We now apply it to the shifted terms of W . [Lemma 3](#), applied to the k -th term with $h = h_k$, gives $\mathbb{E}_{\mathbf{x}}W(Z_{2,1}) = \sum_{k \geq 1} \omega^{-k} \mathbb{E}[h_k(\eta_{k+1}) \mid Z_{2,0} = z]$, in which the k -th term evaluates the $(k + 1)$ -th future shock with the weight and the threshold of the k -th one. Evaluating [\(A.17\)](#) at $u = \eta_{k+1}$ and taking conditional expectations given $Z_{2,0} = z$,

$$\begin{aligned} \mathbb{E}_{\mathbf{x}}W(Z_{2,1}) &\leq (1 + M^{-\kappa}) \sum_{k \geq 1} \omega^{-k} \mathbb{E}[h_{k+1}(\eta_{k+1}) \mid Z_{2,0} = z] \\ &\quad + \sum_{k \geq 1} \omega^{-k} \lambda_0^{\kappa(k+1)} \mathbb{P}(\eta_{k+1} \in I_{k+1} \mid Z_{2,0} = z). \end{aligned}$$

The first sum is a shifted copy of W . Reindexing with $j = k + 1$, and using that $W(z)$ is finite by [Proposition 2](#),

$$\begin{aligned} \sum_{k \geq 1} \omega^{-k} \mathbb{E}[h_{k+1}(\eta_{k+1}) \mid Z_{2,0} = z] &= \omega \sum_{j \geq 2} \omega^{-j} \mathbb{E}[h_j(\eta_j) \mid Z_{2,0} = z] \\ &= \omega W(z) - \mathbb{E}[h_1(\eta_1) \mid Z_{2,0} = z]. \end{aligned}$$

Since $\omega(1 + M^{-\kappa}) = \rho_2$ and $1 + M^{-\kappa} \geq 1$, the first term of the previous bound is at most $\rho_2 W(z) - \mathbb{E}[h_1(\eta_1) \mid Z_{2,0} = z]$. The second sum is controlled by [Lemma A.5](#). Since $I_{k+1} \subseteq [-(1 + M)\lambda_0^{k+1}, (1 + M)\lambda_0^{k+1}]$, [Lemma A.5](#) applied with $k + 1$ in place of k , the inclusion $I_{k+1} \subseteq \{|u| > \lambda_0^k\}$ and [\(A.15\)](#) give

$$\mathbb{P}(\eta_{k+1} \in I_{k+1} \mid Z_{2,0} = z) \leq C_M \mathbb{P}(|\eta_t| > \lambda_0^k) \leq C_M \lambda_0^{-k\gamma} \mathbb{E}|\eta_t|^\gamma,$$

whence

$$\begin{aligned} \sum_{k \geq 1} \omega^{-k} \lambda_0^{\kappa(k+1)} \mathbb{P}(\eta_{k+1} \in I_{k+1} \mid Z_{2,0} = z) &\leq C_M \lambda_0^\kappa \mathbb{E}|\eta_t|^\gamma \sum_{k \geq 1} q^k \\ &= C_M \lambda_0^\kappa \mathbb{E}|\eta_t|^\gamma \frac{q}{1 - q} =: b_2, \end{aligned}$$

a constant independent of z . Inserting these two bounds in the bound on $\mathbb{E}_{\mathbf{x}}W(Z_{2,1})$ gives [\(4.7\)](#). $\square$

Proof of [Theorem 1](#). The proof adds the causal state to [Proposition 3](#), cancels the term (c) of [\(4.8\)](#) through the choice of K , and absorbs the remaining constant into a bounded set. Fix $\rho_2 \in (\omega, 1)$, let b_2 be given by [Proposition 3](#), and let $\mathbf{x} \in \mathbb{R}^p$ and $z = A_2\mathbf{x}$; when $r \geq 1$ write also $z_1 = A_1\mathbf{x}$. By [\(2.5\)](#), $Z_{2,1} = J_2z + a_2\eta_1$ under $\mathbb{P}_{\mathbf{x}}$, and when $r \geq 1$, $Z_{1,1} = J_1z_1 + a_1\eta_1$. We first bound the moment of the shock that enters the causal state. By [\(4.4\)](#), $\mathbb{E}_{\mathbf{x}}|\eta_1|^\kappa = \mathbb{E}[|\eta_1|^\kappa \mid Z_{2,0} = z]$, and [\(A.16\)](#) with $k = 1$ gives $\mathbb{E}_{\mathbf{x}}|\eta_1|^\kappa \leq \mathbb{E}[h_1(\eta_1) \mid Z_{2,0} = z] + \lambda_0^\kappa$, the part of the bound growing with z being exactly the quantity subtracted in [\(4.7\)](#). This yields a one-step bound on PV . Suppose first $r \geq 1$.

[Lemma 2](#) and subadditivity give $\|Z_{1,1}\|_*^\kappa \leq \rho_1^\kappa \|z_1\|_*^\kappa + \|a_1\|_*^\kappa |\eta_1|^\kappa$. Taking expectations and using the previous bound and [Proposition 3](#) yields (4.8), which is well defined since $\mathbb{E}[h_1(\eta_1) \mid Z_{2,0} = z] \leq \omega W(z) < \infty$. If $r = 0$ the causal term is absent and [Proposition 3](#) with $K = 1$ gives directly $PV(\mathbf{x}) \leq 1 + \rho_2 W(z) - \mathbb{E}[h_1(\eta_1) \mid Z_{2,0} = z] + b_2$. If $r \geq 1$ then $K \geq \|a_1\|_*^\kappa$, so term (c) of (4.8) is nonpositive and may be dropped. Let ρ_0 be the slower of the two contraction rates and b_0 the resulting constant,

$$\rho_0 = \begin{cases} \max(\rho_1^\kappa, \rho_2), & r \geq 1, \\ \rho_2, & r = 0, \end{cases} \quad b_0 = \begin{cases} 1 - \rho_0 + \|a_1\|_*^\kappa \lambda_0^\kappa + Kb_2, & r \geq 1, \\ 1 - \rho_0 + b_2, & r = 0. \end{cases}$$

In either case $\rho_0 < 1$ and $PV(\mathbf{x}) \leq \rho_0 V(\mathbf{x}) + b_0$. It remains to absorb the constant. Let $\rho \in (\rho_0, 1)$, let $C_0 \in \mathcal{B}(\mathbb{R}^p)$ be bounded, and $C = \{\mathbf{x} \in \mathbb{R}^p : V(\mathbf{x}) \leq b_0/(\rho - \rho_0)\} \cup C_0$. Off C we have $b_0 < (\rho - \rho_0)V$, hence $PV \leq \rho_0 V + (\rho - \rho_0)V = \rho V$; on C , $PV \leq \rho_0 V + b_0 \leq \rho V + b_0$. Therefore $\Delta V = PV - V \leq -(1 - \rho)V + b_0 \mathbf{1}_C$, which is (3.4) with $b = b_0$. The set $\{V \leq b_0/(\rho - \rho_0)\}$ is closed because V is finite and continuous, W being so by [Proposition 2](#), and bounded because $V(\mathbf{x}) \rightarrow \infty$ as $\|\mathbf{x}\| \rightarrow \infty$. Indeed, (4.5) gives $1 + W(z) \geq C_W^{-1} \|z\|^\kappa$, so, $K \geq 1$ and A^{-1} being invertible,

$$V(\mathbf{x}) \geq \|A_1 \mathbf{x}\|_*^\kappa + C_W^{-1} \|A_2 \mathbf{x}\|^\kappa \quad (r \geq 1), \quad V(\mathbf{x}) \geq C_W^{-1} \|A_2 \mathbf{x}\|^\kappa \quad (r = 0).$$

Hence C is bounded, C_0 being bounded.

Finally, we prove that $\int V d\pi < \infty$. Since (X_t) is strictly stationary, $A_2 \mathbf{X}_0 = Z_{2,0}$ has density f_{Z_2} and $\mathbb{E}W(Z_{2,0}) = \int W f_{Z_2} < \infty$ by (4.6). If $r \geq 1$ then $A_1 \mathbf{X}_0 = Z_{1,0}$ and, by (2.6), subadditivity and $\|J_1^j a_1\|_* \leq \rho_1^j \|a_1\|_*$,

$$\mathbb{E}\|Z_{1,0}\|_*^\kappa \leq \mathbb{E}|\eta_t|^\kappa \sum_{j \geq 0} \|J_1^j a_1\|_*^\kappa \leq \mathbb{E}|\eta_t|^\kappa \frac{\|a_1\|_*^\kappa}{1 - \rho_1^\kappa} < \infty.$$

Hence $\int V d\pi = \mathbb{E}[V(\mathbf{X}_0)] < \infty$ in both cases, without using (3.4). $\square$

A.5 Proofs for [Section 3.3](#)

Proof of [Theorem 2](#). By [Proposition 1](#), $(\mathbf{X}_t)$ is Leb-irreducible and aperiodic. By [Theorem 1](#), $V : \mathbb{R}^p \rightarrow [1, \infty)$ is finite everywhere and satisfies (3.4) on a bounded set C , which is small by [Proposition 1](#), hence petite ([Meyn and Tweedie, 2009](#)). Theorem 15.0.1 of [Meyn and Tweedie \(2009\)](#) then gives the existence of a unique invariant probability, which is the law π of $\mathbf{X}_0$ since π is invariant ([Section 2.2](#)), and of constants $\vartheta > 1$ and $c_V < \infty$ such that (3.5) holds. $\square$

For a signed measure ν of total mass zero on a measurable space $(S, \mathcal{S})$ we write $\|\nu\|_{\text{TV}} = \sup_{B \in \mathcal{S}} |\nu(B)|$ for the total variation distance. On $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$, since $V \geq 1$, it is dominated by the V -norm of [Theorem 2 \(Meyn and Tweedie, 2009\)](#),

$$\|\nu\|_{\text{TV}} \leq \frac{1}{2} \|\nu\|_V, \quad \nu(\mathbb{R}^p) = 0. \quad (\text{A.18})$$

Lemma A.7 ([Davydov, 1973](#)). *Let $(\mathbf{X}_t)_{t \in \mathbb{Z}}$ be a strictly stationary Markov chain on the Polish space $\mathbb{R}^p$, with transition kernel P and invariant law π . Then*

$$\beta_{\mathbf{X}}(n) = \int_{\mathbb{R}^p} \|P^n(\mathbf{x}, \cdot) - \pi\|_{\text{TV}} \pi(d\mathbf{x}), \quad n \geq 1.$$

Proof of [Theorem 3](#). By construction, the chain is strictly stationary with marginal law π , and $\mathbb{R}^p$ is Polish, so [Lemma A.7](#) applies to $(\mathbf{X}_t)$. Multiplying the identity of [Lemma A.7](#) by ϑ^n , summing over $n \geq 1$, exchanging the sum and the integral, and using [\(A.18\)](#) and then [\(3.5\)](#),

$$\sum_{n \geq 1} \vartheta^n \beta_{\mathbf{X}}(n) \leq \frac{1}{2} \int_{\mathbb{R}^p} \sum_{n \geq 1} \vartheta^n \|P^n(\mathbf{x}, \cdot) - \pi\|_V \pi(d\mathbf{x}) \leq \frac{c_V}{2} \int_{\mathbb{R}^p} V d\pi,$$

which is finite. This is [\(3.7\)](#). The terms of the sum being nonnegative, each is bounded by the total, which is the second inequality of [\(3.6\)](#).

Since X_t is the first coordinate of $\mathbf{X}_t = (X_t, \dots, X_{t-p+1})'$, we have $\sigma(X_t : t \leq 0) \subseteq \sigma(\mathbf{X}_t : t \leq 0)$ and $\sigma(X_t : t \geq n) \subseteq \sigma(\mathbf{X}_t : t \geq n)$, so the monotonicity of β gives $\beta_X(n) \leq \beta_{\mathbf{X}}(n)$, the first inequality of [\(3.6\)](#). Finally, With $\varsigma = \vartheta^{-1} \in (0, 1)$ and $c_\beta = \frac{c_V}{2} \int_{\mathbb{R}^p} V d\pi < \infty$, [\(3.6\)](#) reads $\beta_X(n) \leq \beta_{\mathbf{X}}(n) \leq c_\beta \varsigma^n$. $\square$

A.6 Proofs for [Section 5](#)

Proof of [Proposition 4](#). We bound the exponential moment $\mathbb{E}_{\mathbf{x}}[\rho^{-\tau_C}]$ by a supermartingale argument, and then apply the Markov and Jensen inequalities. By [\(3.4\)](#), $PV \leq \rho V$ off C and $PV \leq \rho V + b$ everywhere. Let $\tau_C^0 = \inf\{t \geq 0 : \mathbf{X}_t \in C\}$ and fix $\mathbf{y} \notin C$. Under $\mathbb{P}_{\mathbf{y}}$, let $\mathcal{G}_t = \sigma(\mathbf{X}_0, \dots, \mathbf{X}_t)$ and $M_t = \rho^{-(t \wedge \tau_C^0)} V(\mathbf{X}_{t \wedge \tau_C^0})$. On $\{\tau_C^0 > t\}$ we have $\mathbf{X}_t \notin C$, so $\mathbb{E}_{\mathbf{y}}[M_{t+1} \mid \mathcal{G}_t] = \rho^{-(t+1)} PV(\mathbf{X}_t) \leq \rho^{-t} V(\mathbf{X}_t) = M_t$, and on $\{\tau_C^0 \leq t\}$ we have $M_{t+1} = M_t$. Hence (M_t) is a nonnegative supermartingale and, since $V \geq 1$, $\mathbb{E}_{\mathbf{y}}[\rho^{-(t \wedge \tau_C^0)}] \leq \mathbb{E}_{\mathbf{y}}[M_t] \leq V(\mathbf{y})$. Letting $t \rightarrow \infty$ gives $\mathbb{E}_{\mathbf{y}}[\rho^{-\tau_C^0}] \leq V(\mathbf{y})$, and in particular $\tau_C^0 < \infty$ $\mathbb{P}_{\mathbf{y}}$ -almost surely. This bound also holds for $\mathbf{y} \in C$, where $\tau_C^0 = 0$ and $V \geq 1$.

Now $\tau_C = 1 + \inf\{t \geq 0 : \mathbf{X}_{t+1} \in C\}$ is one plus the hitting time τ_C^0 of the chain $(\mathbf{X}_{t+1})_{t \geq 0}$, which starts from $\mathbf{X}_1$, so by the Markov property

$$\mathbb{E}_{\mathbf{x}}[\rho^{-\tau_C}] = \rho^{-1} \mathbb{E}_{\mathbf{x}}[\mathbb{E}_{\mathbf{X}_1}[\rho^{-\tau_C^0}]] \leq \rho^{-1} PV(\mathbf{x}) \leq \rho^{-1}(\rho V(\mathbf{x}) + b),$$

that is, $\mathbb{E}_{\mathbf{x}}[\rho^{-\tau_C}] \leq V(\mathbf{x}) + b/\rho$. Markov's inequality gives

$$\mathbb{P}_{\mathbf{x}}(\tau_C > n) = \mathbb{P}_{\mathbf{x}}(\rho^{-\tau_C} \geq \rho^{-(n+1)}) \leq \rho^{n+1} \mathbb{E}_{\mathbf{x}}[\rho^{-\tau_C}],$$

which is the first bound of (5.1), and Jensen's inequality applied to the convex function $t \mapsto \rho^{-t}$ gives $\rho^{-\mathbb{E}_{\mathbf{x}} \tau_C} \leq \mathbb{E}_{\mathbf{x}}[\rho^{-\tau_C}]$, which is the second. $\square$

A.7 Proofs for Section 6

We keep the notation $Y_{\mathcal{S}} = \sum_{j \in \mathcal{S}} d_j \eta_j$ of Section A.3.3, now with $d_j \in \mathbb{R}^{n_2 \times m}$, and set $\mathcal{I}_n = \{n+1, \dots, n+n_2\}$.

Lemma A.8. *Under Assumption 4(ι), $D_n = [d_{n+1}, \dots, d_{n+n_2}] \in \mathbb{R}^{n_2 \times mn_2}$ has rank n_2 for every $n \geq 0$.*

Proof of Lemma A.8. Since $\Psi_c(V'_1, \dots, V'_p)' = ((\sum_j \Phi_j V_j)', V'_1, \dots, V'_{p-1})'$ for $V_j \in \mathbb{R}^{m \times m}$, the $(k+1)$ -th block of $\Psi_c^k E$ is I_m and its later blocks vanish, $0 \leq k \leq p-1$. The matrix $[E, \Psi_c E, \dots, \Psi_c^{p-1} E]$ is thus block upper triangular with diagonal blocks I_m , hence invertible, as in the proof of Lemma 1 for $m=1$. As in that proof, $A_2 \Psi_c^k = J_2^k A_2$, and A_2 has rank n_2 , so $[a_2, J_2 a_2, \dots, J_2^{p-1} a_2]$ has rank n_2 . By the Cayley–Hamilton theorem, so has $[a_2, J_2 a_2, \dots, J_2^{n_2-1} a_2]$, and so has $D_n = -J_2^{-n-n_2} [J_2^{n_2-1} a_2, \dots, a_2]$, J_2 being invertible. $\square$

Lemma A.9. *Suppose Assumption 4 holds, let $n \geq 0$, $\mathcal{I}_n \subseteq \mathcal{S} \subseteq \{1, 2, \dots\}$ and $D_n^+ = D_n' (D_n D_n')^{-1}$. Then $Y_{\mathcal{S}}$ has a density $f_{\mathcal{S}}$ which is positive, continuous and bounded by $H_g(\|D_n^+\|)^{n_2} / \text{Leb}(B_1)$, with B_1 the unit ball of $\mathbb{R}^{n_2}$, and satisfies $f_{\mathcal{S}}(w) \leq H_g(\|D_n^+\| \|y - w\|)^{n_2} f_{\mathcal{S}}(y)$ for $y, w \in \mathbb{R}^{n_2}$. Consequently, Lemmas 1 and A.4 hold with $\mathbb{R}^{n_2}$, g and $\int_{\mathbb{R}^m}$ in place of $\mathbb{R}^s$, f_{η} and $\int_{\mathbb{R}}$, and with $C_R = \max_{n \in \{0, n_2\}} H_g(\|D_n^+\|)^{n_2} / \text{Leb}(B_1)$ and $H(\delta) = \max_{n \in \{0, n_2\}} H_g(\|D_n^+\| \delta)^{n_2}$.*

Proof of Lemma A.9. Only the density of the block needs a new argument, because D_n is no longer square. We work with whole shocks $\eta_j \in \mathbb{R}^m$, whose components may be dependent. The block $\xi = (\eta'_{n+1}, \dots, \eta'_{n+n_2})' \in \mathbb{R}^N$, $N = mn_2$, has density $f_{\xi}(x) = \prod_{j=1}^{n_2} g(x_j)$, $x = (x'_1, \dots, x'_{n_2})'$, $x_j \in \mathbb{R}^m$. Each block of a vector $\Delta \in \mathbb{R}^N$ has norm at most $\|\Delta\|$, so Assumption 4(ι) gives

$$f_{\xi}(x + \Delta) \leq H_g(\|\Delta\|)^{n_2} f_{\xi}(x), \quad x, \Delta \in \mathbb{R}^N. \quad (\text{A.19})$$

Write $D = D_n$, which has rank n_2 by Lemma A.8, so that $DD^+ = I_{n_2}$, and let $U \in \mathbb{R}^{N \times (N-n_2)}$ have orthonormal columns spanning $\ker D$ (U is absent if $N = n_2$). The matrix $\bar{D} = (D', U)'$ satisfies $\bar{D} \bar{D}' = \text{diag}(DD', I)$ and $\bar{D}^{-1}(w', v')' = D^+ w + Uv$, since

$DU = 0$. By the change of variables $x \mapsto \bar{D}x$, $D\xi$ has density

$$f_{D\xi}(w) = \det(DD')^{-1/2} \int_{\mathbb{R}^{N-n_2}} f_\xi(D^+w + Uv) dv \in [0, \infty], \quad w \in \mathbb{R}^{n_2}.$$

Since $D^+w + Uv = (D^+y + Uv) + D^+(w - y)$, (A.19) gives $f_{D\xi}(w) \leq H_g(\|D^+\| \|w - y\|)^{n_2} f_{D\xi}(y)$, and averaging over y in the unit ball centered at w gives $f_{D\xi}(w) \leq H_g(\|D^+\|)^{n_2} / \text{Leb}(B_1)$. The integrand is positive by Assumption 4(ι). If $w_k \rightarrow w$ with $\|w_k - w\| \leq 1$, the integrands converge pointwise, g being continuous, and are dominated by $H_g(\|D^+\|)^{n_2} f_\xi(D^+w + Uv)$, integrable since $f_{D\xi}(w) < \infty$; so $f_{D\xi}$ is continuous. The rest of the proof of Lemma A.3, which writes $Y_S = D_n \xi + Y_{\mathcal{T}}$ with $\mathcal{T} = \mathcal{S} \setminus \mathcal{I}_n$ and $f_S(y) = \mathbb{E}[f_{D_n \xi}(y - Y_{\mathcal{T}})]$, transfers these properties from $f_{D_n \xi}$ to f_S verbatim. Finally, the index sets of $Z_{2,0}$ and of R_k contain $\mathcal{I}_0$, except for R_k with $k \leq n_2$, whose index set contains $\mathcal{I}_{n_2}$; this gives Lemma 1 and (A.10), and the proofs of (A.11)–(A.13) apply verbatim, with $u \in \mathbb{R}^m$. $\square$

Proof of Theorem 4. By Assumption 4(ι) and (ιv), Brockwell et al. (2012), applied to the state equation $\mathbf{Y}_t = \Psi_c \mathbf{Y}_{t-1} + E \eta_t$, whose strictly stationary solutions are the states of those of (6.1), shows that (6.1) has a unique strictly stationary solution, with state $\mathbf{Y}_t = A Z_t$, where $Z_{1,t} = \sum_{j \geq 0} J_1^j a_1 \eta_{t-j}$ and $Z_{2,t} = \sum_{j \geq 1} d_j \eta_{t+j}$ converge absolutely almost surely. By Lemma A.9, the noncausal state $Z_{2,t}$ has a positive, continuous and bounded density f_{Z_2} on $\mathbb{R}^{n_2}$. The predictive density of Y_t given $\mathbf{Y}_{t-1} = \mathbf{y}$ is therefore defined at every $\mathbf{y} \in \mathbb{R}^{mp}$, and is given by (2.7) (Gouriéroux and Jasiak, 2026), with the notational changes listed below. Hence $(\mathbf{Y}_t)$ is a Markov chain on $\mathbb{R}^{mp}$ whose transition kernel P is defined by this density as in Section 2.2. Lemma A.2 and (A.7) and (A.8) are replaced by Assumption 4($\iota \iota \iota$), which also makes g bounded, since averaging $g(u) \leq H_g(1)g(v)$ over v in the unit ball of $\mathbb{R}^m$ centered at u bounds $g(u)$ by $H_g(1)$ divided by the volume of that ball. All other proofs of Sections 2 to 4 apply with notational changes only: x, u, v, w_k and the shocks take values in $\mathbb{R}^m$, $|\eta_j|$ becomes $\|\eta_j\|$, f_η becomes g , $\theta' \mathbf{y}$ becomes $E' \Psi_c \mathbf{y}$, s, p and α become n_2, mp and $\alpha_0, r \geq 1$ becomes $n_1 \geq 1$, the window of Lemma A.5 becomes a ball, X_t becomes the first block Y_t of $\mathbf{Y}_t$, and λ_ϕ^{-1} becomes the spectral radius of J_1 . Indeed, apart from the densities, these proofs use the shocks only through their norms and the identities $Z_{2,0} = d_k \eta_k + R_k$, $Z_{2,1} = J_2 Z_{2,0} + a_2 \eta_1$, $d_k = J_2 d_{k+1}$ and $J_2 d_1 = -a_2$. They never use the invertibility of J_1 , so Lemma 2 holds with operator norms even when J_1 is singular. $\square$

Proof of Corollary 1. Since $\psi^*(B) = -\psi_s^{-1} B^s \psi(F)$, the first row of (6.3) is (6.2) multiplied by $-\psi_s^{-1} B^s$, which $-\psi_s F^s$ undoes, so both systems have the same solutions. The polynomial $\mathcal{A}$ on the left of (6.3) has degree $\bar{p}$, as $s \leq p$, and $\mathcal{A}(0) = I$, as $s \geq 1$. It is block triangular, so, by the identity $\det(I - \Psi_c z) = \det \mathcal{A}(z)$, the nonzero eigenvalues of its companion matrix are the reciprocals of the roots of ϕ and ψ^* and the nonzero

eigenvalues of the companion matrix of Φ_x ; [Assumption 4](#)(ι) holds with $n_2 \geq s$. The shocks $(-\psi_s^{-1}\varepsilon_{t-s}, \varepsilon'_{x,t})'$ at distinct dates are functions of disjoint, hence independent, families, so they are *i.i.d.*, with density $f_\eta(u_1)g_x(u_2)$, where g_x is the density of $\varepsilon_{x,t}$. Its local ratio bound at distance δ is at most $H_\eta(\delta)H_{g_x}(\delta)$, and it has a finite moment of order α_0 , so [Theorem 4](#) applies.

For the MARDL model, applying $-\psi_s^{-1}B^s$ to (6.4) gives $\theta(B)y_t + \psi_s^{-1}\sum_{i=0}^{k_x}\beta'_i x_{t-i} = \eta_t$, which $-\psi_s F^s$ undoes. Hence $Y_t = (y_t, x'_t)'$ satisfies $\tilde{\mathcal{A}}(B)Y_t = (\eta_t, \varepsilon'_{x,t})'$, where $\tilde{\mathcal{A}}$ is block triangular with diagonal blocks $\theta(B)$ and $\Phi_x(B)$, has degree $\max(p, p_x, k_x)$ and determinant $\theta(z)\det\Phi_x(z)$, and $\tilde{\mathcal{A}}(0)$ has first row $(1, \mathbf{c}')$ with $\mathbf{c} = \psi_s^{-1}\beta_0$. Multiplying by $\tilde{\mathcal{A}}(0)^{-1}$, of determinant one, gives a VAR of the form (6.1) with the same determinant and *i.i.d.* shocks $(\eta_t - \mathbf{c}'\varepsilon_{x,t}, \varepsilon'_{x,t})'$, of density $f_\eta(u_1 + \mathbf{c}'u_2)g_x(u_2)$. This density has local ratio bound at most $H_\eta((1 + \|\mathbf{c}\|)\delta)H_{g_x}(\delta)$ and a finite moment of order α_0 , so the argument above applies. $\square$

Proof of Corollary 2. (a) Polynomials in B and $F = B^{-1}$ with scalar coefficients commute, so applying $v(F)\zeta(B)$ to (2.1) shows that $\tilde{X}_t$ solves (6.5), and $\tilde{X}_t$ is strictly stationary, as the same linear function of $(X_{t-r'}, \dots, X_{t+s'})$ for every t . Multiplying by $-\psi_s^{-1}B^s$, (6.5) is equivalent to $\theta(B)\tilde{X}_t = w_t$, $w_t = -\psi_s^{-1}v(F)\zeta(B)\varepsilon_{t-s}$, an autoregressive moving average (ARMA) equation driven by the *i.i.d.* sequence $(\varepsilon_{t-s+s'})$. As θ has no root on the unit circle, its strictly stationary solution is unique ([Brockwell et al., 2012](#)). Since $\sigma(\tilde{X}_t : t \leq 0) \subseteq \sigma(X_t : t \leq s')$ and $\sigma(\tilde{X}_t : t \geq n) \subseteq \sigma(X_t : t \geq n - r')$, monotonicity and stationarity give $\beta_{\tilde{X}}(n) \leq \beta_X(n - r' - s')$, and [Theorem 3](#) concludes.

(b) The VAR $\det\Phi(B)\mathcal{Y}_t = \eta_t$ is of the form (6.1), and its companion matrix $\check{\Psi}_c$ satisfies $\det(I - \check{\Psi}_c z) = (\det\Phi(z))^m$, so [Theorem 4](#) applies to it. As $\Phi \operatorname{adj}\Phi = (\det\Phi)I_m$ and $\det\Phi(B)$ commutes with $\Lambda(B)$, $Y_t = \operatorname{adj}\Phi(B)\Lambda(B)\mathcal{Y}_t$ is strictly stationary and solves $\Phi(B)Y_t = \Lambda(B)\eta_t$, uniquely by [Brockwell et al. \(2012\)](#), as $\det\Phi$ has no root on the unit circle, and, as in (a), $\beta_Y(n) \leq \beta_{\mathcal{Y}}(n - k)$ for $n > k = \deg(\operatorname{adj}\Phi\Lambda)$, since Y_t is a function of $\mathcal{Y}_{t-k}, \dots, \mathcal{Y}_t$. $\square$