Anders Rahbek
Rasmus Søndergaard Pedersen
University of Copenhagen

September 2024

Part V

SREs: Stationarity, Ergodicity, Tails and Limit Theory

V.1 Stochastic Recurrence Equations

In the previous chapters, we considered the stochastic properties of Markov chains $\{X_t\}_{t=0,1,...}$. It was shown that geometric ergodicity implies that X_0 can be assigned a particular distribution, $X_0 \overset{D}{=} X_0^*$ such that the resulting process $\{X_t^*\}_{t=0,1,...}$ is stationary. Moreover, the LLN in Theorem I.4.2 stated that for any initial value of X_0, $T^{-1}\sum_{t=1}^{T} X_t \xrightarrow{p} \mathbb{E}[X_t^*]$ provided that $\mathbb{E}[|X_t^*|] < \infty$. That is, under geometric ergodicity the initial value of the Markov chain plays no role for the stochastic limit results. An important consequence of the geometric ergodicity is that $\{X_t^*\}_{t=0,1,...}$ is so-called ergodic (which we define below). In particular, there are LLNs and CLTs that apply to stationary and ergodic processes parallel to Theorems I.4.2 and I.4.4.

In this chapter we consider a specific class of Markov chains for $X_t \in \mathbb{R}^d$ given by

$$X_t = A_t X_{t-1} + B_t,$$

for some initial value X_0, where A_t is a $d \times d$ random matrix and B_t is a d-dimensional random vector, and $\{(A_t, B_t)\}_{t=1,2,...}$ is an i.i.d. process such that (A_t, B_t) and $\{X_{t-1}, X_{t-2}, \ldots, X_0\}$ are independent for all $t \geq 1$. The above equation for X_t is a so-called *stochastic recurrence equation* (SRE), and, as illustrated later in Section V.3, this class of processes covers a wide range of processes applied in financial econometric modelling. The aim of the chapter is to provide conditions such that the stationary version, $\{X_t^*\}_{t=0,1,...}$, of $\{X_t\}_{t=0,1,...}$ exists, with $X_t^* = A_t X_{t-1}^* + B_t$. Moreover, we present an *explicit expression* for the stationary solution to the SRE, X_t^*, and provide conditions ensuring finite moments and ergodicity such that a LLN and a CLT apply to $\{X_t^*\}_{t=0,1,...}$. The conditions for stationarity and finite moments are sharp in the sense that they are essentially necessary. Moreover, the stated conditions enable us to characterize the tail shape of the unconditional distribution of X_t^*. Importantly, under mild conditions the distribution is *heavy-tailed* which is of particular relevance in actuarial sciences and risk management and mimics the tails of the empirical distribution of, say, equity returns observed in practice.

Most of the results stated in the chapter can be found in the recent textbook by Burazcewski et al. (2016) [BDM henceforth] that provides a comprehensive treatment of SREs as well as detailed proofs of the stated results. We emphasize that conditions ensuring geometric ergodicity of $\{X_t\}_{t=0,1,\ldots}$ can be derived by applying the drift criterion; cf. Part I and BDM (Section 2.2). The conditions for stationarity and finite moments provided in this chapter are – in addition to provide further knowledge about X_t^* – typically milder as they, for instance, do not require the Markov chain to have a positive, continuous transition density.

Our main focus is to consider cases where the SRE process is stationary, and we state limit results only in terms of the stationary version of the process. In order to ease the notation, throughout, whenever a given process is stationary we write it in terms of X_t and not X_t^*. Moreover, as is standard in the literature, we let stationary processes run over all integers $\mathbb{Z}$ instead of $\{0, 1, \ldots\}$.[1] Consequently, we pay attention to processes $\{X_t\}_{t\in\mathbb{Z}}$ of the form

$$X_t = A_t X_{t-1} + B_t, \quad t \in \mathbb{Z}, \tag{V.1}$$

with $X_t \in \mathbb{Z}^d$ and with A_t and B_t given as above. It is assumed that $\{(A_t, B_t)\}_{t\in\mathbb{Z}}$ is an i.i.d. process and (A_t, B_t) and $\{X_{t-1}, X_{t-2}, \ldots\}$ are independent for all $t \in \mathbb{Z}$. We say that X_t obeys an SRE, if it satisfies (V.1).

V.2 Ergodicity and Limit Theory

In this section we define the notion of ergodicity and present limit theorems for stationary and ergodic processes. These results will be used throughout in the remainder of the course when considering statistical inference in models for time-varying conditional volatility, such as ARCH. We emphasize that there are different ways of defining ergodicity. We say that a stationary process $\{X_t\}_{t\in\mathbb{Z}}$, with $X_t \in \mathbb{R}^d$, is ergodic if and only if for every measurable function $f : (\mathbb{R}^d)^\infty \to \mathbb{R}$ with $\mathbb{E}[|f(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots)|] < \infty$, as $T \to \infty$

$$\frac{1}{T}\sum_{t=1}^{T} f(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots) \overset{a.s.}{\to} \mathbb{E}\left[f(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots)\right], \tag{V.2}$$

[1] This is without loss of generality: The process $\{X_t^*\}_{t=0,1,\ldots}$ is stationary if and only if, there exists a stationary process $\{Y_t\}_{t\in\mathbb{Z}}$ satisfying $\{X_t^*\}_{t=0,1,\ldots} \overset{D}{=} \{Y_t\}_{t=0,1,\ldots}$.

where $\overset{a.s.}{\to}$ denotes almost sure convergence[2]. By definition, ergodicity ensures that a (strong) LLN applies to the stationary process $\{X_t\}_{t\in\mathbb{Z}}$. This parallels geometric ergodicity of Markov chains that ensures that a LLN holds (Theorem I.4.2). Recall that geometric ergodicity implies the "mixing" property that X_t and X_{t+h} are independent as $h \to \infty$. Likewise, ergodicity implicitly ensures that observations cannot be "too dependent". To see this, consider the process $\{X_t\}_{t\in\mathbb{Z}}$, with $X_t \in \mathbb{R}$ such that $\mathbb{P}(X_t = Z) = 1$ for some non-degenerate random variable Z with $\mathbb{E}[|Z|] < \infty$. Clearly the process is stationary, but it is not ergodic, as the average of the perfectly dependent process $T^{-1}\sum_{t=1}^{T} X_t = Z$ (almost surely) does not converge to $\mathbb{E}[X_t] = \mathbb{E}[Z]$.

As mentioned in the introduction, the stationary version of a geometric ergodic process is ergodic:

Corollary V.2.1 *Suppose that a Markov chain $\{Y_t\}_{t=0,1,\ldots}$ satisfies the drift criterion in Part I such that it is geometrically ergodic. Then there exists a stationary and ergodic process $\{X_t\}_{t\in\mathbb{Z}}$ with $\{X_t\}_{t=0,1,\ldots} \overset{d}{=} \{Y_t^*\}_{t=0,1,\ldots}$.*

Before presenting a LLN and CLT for stationary and ergodic processes, we state the following result that any nice deterministic function of a stationary and ergodic process yields itself a stationary and ergodic process:

Theorem V.2.1 *With $X_t \in \mathbb{R}^d$, let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic process. For some measurable $f : (\mathbb{R}^d)^\infty \to \mathbb{R}$, let*

$$Y_t = f(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots), \quad t \in \mathbb{Z}.$$

Suppose that Y_t is finite almost surely (e.g., $\mathbb{E}[|Y_t|] < \infty$) for some t. Then the process $\{Y_t\}_{t\in\mathbb{Z}}$ is stationary and ergodic.

Note that by definition of ergodicity, (V.2) ensures that a (strong) LLN applies. For convenience, we state this as a theorem, typically referred to as the Ergodic Theorem.[3]

Theorem V.2.2 *With $X_t \in \mathbb{R}^d$, let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic process. For some measurable $f : (\mathbb{R}^d)^\infty \to \mathbb{R}$, let $Y_t = f(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots)$, $t \in \mathbb{Z}$. Suppose that $\mathbb{E}[|Y_t|] < \infty$. Then as $T \to \infty$,*

$$\frac{1}{T}\sum_{t=1}^{T} Y_t \overset{p}{\to} \mathbb{E}[Y_t].$$

[2] That is, $P\left(\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} f(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots) = \mathbb{E}[f(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots)]\right) = 1$. You may recall that almost sure convergence implies convergence in probability.

[3] Note that by definition of ergodicity, the convergence holds almost surely. We state the weaker result that the average converges with probability approaching one, as we nowhere need almost sure convergence when considering the stochastic properties of estimators later on.

We also have the following CLT for martingale differences.

Theorem V.2.3 *With $X_t \in \mathbb{R}^d$, let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic process. For some measurable $f : (\mathbb{R}^d)^\infty \to \mathbb{R}$, let*

$$Y_t = f(\dots, X_{t-1}, X_t, X_{t+1}, \dots), \quad t \in \mathbb{Z},$$

and let $\mathcal{F}_t$ denote the natural filtration generated by $\{X_s\}_{s\le t}$. Assume that $\mathbb{E}[Y_t|\mathcal{F}_{t-1}] = 0$ a.s. and $0 < \mathbb{E}[Y_t^2] < \infty$. Then as $T \to \infty$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_t \xrightarrow{D} N(0, \mathbb{E}[Y_t^2]).$$

The result is essentially a corollary to Theorem I.4.4. In particular with $Y_t = X_t$, we see that the conditions in that theorem are easily checked:

$$\text{(i)} \quad \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[X_t^2|\mathcal{F}_{t-1}] \xrightarrow{p} \mathbb{E}[\mathbb{E}[X_t^2|\mathcal{F}_{t-1}]] = \mathbb{E}[X_t^2] > 0,$$

where we have used that $\mathbb{E}[\mathbb{E}[X_t^2|\mathcal{F}_{t-1}]] < \infty$ and Theorem V.2.2. Likewise,

$$\begin{aligned}\text{(ii)} \quad \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[X_t^2 \mathbb{I}(|X_t| > \delta\sqrt{T})\right] &= \mathbb{E}[X_t^2 \mathbb{I}(|X_t| > \delta\sqrt{T})] \\ &= \int x^2 \mathbb{I}(|x| > \delta\sqrt{T}) dP \to 0 \quad \text{as } T \to \infty,\end{aligned}$$

where we have used that $\mathbb{I}(|x| > \delta\sqrt{T}) \to 0$, $\mathbb{E}[X_t^2] = \int x^2 dP < \infty$ as well as dominated convergence.

Remark V.2.1 *Ergodicity does not ensure that a CLT for non-martingale differences, such as Theorem I.4.3, holds. In order to have such a result, one would typically rely on showing that the process is geometrically ergodic by relying on the drift criterion.*

V.3 Examples

The following examples illustrate that many time series processes within financial econometrics belongs to the class of SREs.

Example V.3.1 (AR(1)) *Consider the AR(1) process from Part I with*

$$x_t = \rho x_{t-1} + \varepsilon_t,$$

with $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ *and i.i.d. process with* $\mathbb{E}[\varepsilon_t] = 0$ *and* $\mathbb{V}[\varepsilon_t] = \sigma^2$ *with* $0 < \sigma^2 < \infty$. *We see that this process is given by and SRE with* $d = 1$, $A_t = \rho$ *(constant) and* $B_t = \varepsilon_t$.

Example V.3.2 (VAR(1)) *Consider the VAR(1) from Part I with*

$$X_t = AX_{t-1} + \varepsilon_t,$$

with $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ *an i.i.d. process with* $\mathbb{E}[\varepsilon_t] = 0$ *and* $\mathbb{V}[\varepsilon_t] = \Omega$ *positive definite. The process obeys an SRE with* $A_t = A$ *and* $B_t = \varepsilon_t$.

Example V.3.3 (ARCH(1)) *Consider Engle's (1982) ARCH(1) process from Part I given by*

$$\begin{aligned} x_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega + \alpha x_{t-1}^2, \end{aligned}$$

with $\alpha \geq 0$, $\omega > 0$,*and* $\{Z_t\}_{t\in\mathbb{Z}}$ *an i.i.d. process with* $Z_t \overset{D}{=} N(0,1)$. *Recall that (almost surely)* $\mathbb{E}[x_t|x_{t-1}] = 0$ *and* $\mathbb{E}[x_t^2|x_{t-1}] = \sigma_t^2$. *In fact,*

$$x_t|x_{t-1} \overset{D}{=} N(0, \sigma_t^2).$$

Consider the random vector

$$\begin{pmatrix} A_t \\ B_t \end{pmatrix} \overset{D}{=} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \alpha & 0 \\ 0 & \omega \end{bmatrix}\right),$$

Then with $X_t = x_t \in \mathbb{R}$ *satisfying (V.1), we have that*

$$x_t|x_{t-1} \overset{D}{=} N(\mathbb{E}[x_t|x_{t-1}], \mathbb{V}[x_t|x_{t-1}]),$$

with (almost surely)

$$\begin{aligned} \mathbb{E}[x_t|x_{t-1}] &= \mathbb{E}[A_t x_{t-1} + B_t|x_{t-1}] \\ &= \mathbb{E}[A_t|x_{t-1}]x_{t-1} + \mathbb{E}[B_t|x_{t-1}] \\ &= 0x_{t-1} + 0 \\ &= 0, \end{aligned}$$

and

$$
\begin{aligned}
\mathbb{V}[x_t|x_{t-1}] &= \mathbb{E}[x_t^2|x_{t-1}] \\
&= \mathbb{E}[(A_t x_{t-1} + B_t)^2|x_{t-1}] \\
&= \mathbb{E}[A_t^2|x_{t-1}]x_{t-1}^2 + \mathbb{E}[B_t^2|x_{t-1}] + 2\mathbb{E}[A_t B_t|x_{t-1}]x_{t-1} \\
&= \alpha x_{t-1}^2 + \omega + 2 \times 0 \times x_{t-1} \\
&= \omega + \alpha x_{t-1}^2.
\end{aligned}
$$

We conclude that the ARCH(1) process as an SRE representation. Consequently the stochastic properties of the ARCH(1) process, such as stationarity, ergodicity and finite moments, can be derived by making use of the SRE representation.

Example V.3.4 (BEKK) *Consider the $(d+1)$-dimensional vector*

$$
\begin{pmatrix} m_t \\ B_t \end{pmatrix} \overset{D}{=} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & \Omega \end{bmatrix} \right),
$$

with Ω constant and positive definite $(d \times d)$ matrix. For a constant $(d \times d)$ matrix A, let $A_t = m_t A$. Then with $X_t \in \mathbb{R}^d$ given by (V.1),

$$
X_t|X_{t-1} \overset{D}{=} N(\mathbb{E}[X_t|X_{t-1}], \mathbb{V}[X_t|X_{t-1}]),
$$

with (almost surely)

$$
\begin{aligned}
\mathbb{E}[X_t|X_{t-1}] &= \mathbb{E}[A_t X_{t-1} + B_t|X_{t-1}] \\
&= \mathbb{E}[A_t|X_{t-1}]X_{t-1} + \mathbb{E}[B_t|X_{t-1}] \\
&= 0AX_{t-1} + 0_{d\times 1} \\
&= 0_{d\times 1},
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{V}[X_t|X_{t-1}] \\
&= \mathbb{E}[X_t X_t'|X_{t-1}] \\
&= \mathbb{E}[(A_t X_{t-1} + B_t)(A_t X_{t-1} + B_t)'|X_{t-1}] \\
&= \mathbb{E}[A_t X_{t-1} X_{t-1}' A_t'|X_{t-1}] + \mathbb{E}[B_t B_t'|X_{t-1}] + \mathbb{E}[A_t X_{t-1} B_t' + B_t X_{t-1}' A_t'|X_{t-1}] \\
&= \mathbb{E}[m_t^2|X_{t-1}]AX_{t-1}X_{t-1}'A' + \mathbb{E}[B_t B_t'|X_{t-1}] + AX_{t-1}\mathbb{E}[m_t B_t'|X_{t-1}] + \mathbb{E}[B_t m_t|X_{t-1}](AX_{t-1})' \\
&= \mathbb{E}[m_t^2]AX_{t-1}X_{t-1}'A' + \mathbb{E}[B_t B_t'] + AX_{t-1}\mathbb{E}[m_t B_t'] + \mathbb{E}[B_t m_t](AX_{t-1})' \\
&= AX_{t-1}X_{t-1}'A' + \Omega.
\end{aligned}
$$

Similar to the ARCH(1) example given above, we may alternatively write the process as

$$X_t = \Omega_t^{1/2} Z_t$$
$$\Omega_t = \Omega + A X_{t-1} X_{t-1}' A,$$

with $\{Z_t\}_{t\in\mathbb{Z}}$ *an i.i.d. process with* $Z_t \overset{D}{=} N(0, I_d)$ *and* Z_t *independent of* $\{X_{t-1}, X_{t-2}, \dots\}$*, and* $\Omega_t^{1/2}$ *is the (symmetric) square-root of* Ω_t*. Here* X_t *is given in terms of a multivariate ARCH process with a so-called Baba-Engle-Kraft-Kroner (BEKK) formulation of the conditional covariance matrix* $\mathbb{V}[X_t|X_{t-1}]$*, as originally considered in Engle and Kroner (1995). We return to this type of process later on when considering modelling of time varying conditional covariance matrices as used within dynamic allocation of assets.*

Example V.3.5 (DAR) *Consider the bivariate vector*

$$\begin{pmatrix} A_t \\ B_t \end{pmatrix} \overset{D}{=} N\left(\begin{bmatrix} \phi \\ 0 \end{bmatrix}, \begin{bmatrix} \alpha & 0 \\ 0 & \omega \end{bmatrix}\right),$$

with constants $\phi \in \mathbb{R}$*,* $\alpha \geq 0$*,* $\omega > 0$*. Then with* $X_t = x_t \in \mathbb{R}$ *satisfying (V.1), we have that*

$$x_t|x_{t-1} \overset{D}{=} N(\mathbb{E}[x_t|x_{t-1}], \mathbb{V}[x_t|x_{t-1}]),$$

with (almost surely),

$$\begin{aligned} \mathbb{E}[x_t|x_{t-1}] &= \mathbb{E}[A_t x_{t-1} + B_t|x_{t-1}] \\ &= \mathbb{E}[A_t|x_{t-1}]x_{t-1} + \mathbb{E}[B_t|x_{t-1}] \\ &= \mathbb{E}[A_t]x_{t-1} + 0 \\ &= \phi x_{t-1}, \end{aligned}$$

and

$$\begin{aligned} &\mathbb{V}[x_t|x_{t-1}] \\ &= \mathbb{E}[x_t^2|x_{t-1}] - (\mathbb{E}[x_t|x_{t-1}])^2 \\ &= \mathbb{E}[(A_t x_{t-1} + B_t)^2|x_{t-1}] - (\phi x_{t-1})^2 \\ &= \mathbb{E}[A_t^2 x_{t-1}^2|X_{t-1}] + \mathbb{E}[B_t^2|x_{t-1}] + \mathbb{E}[A_t B_t x_{t-1}|X_{t-1}] - (\phi x_{t-1})^2 \\ &= \mathbb{E}[A_t^2]x_{t-1}^2 + \mathbb{E}[B_t^2] + 0 - (\phi x_{t-1})^2 \\ &= (\alpha + \phi^2)x_{t-1}^2 + \omega - (\phi x_{t-1})^2 \\ &= \omega + \alpha x_{t-1}^2. \end{aligned}$$

Here X_t is given in terms of a so-called double autoregressive (DAR) process of order one with Gaussian innovations. An alternative formulation of X_t is

$$x_t = \phi x_{t-1} + \sigma_t z_t,$$
$$\sigma_t^2 = \omega + \alpha x_{t-1}^2,$$

where $\{z_t\}_{t\in\mathbb{Z}}$ is an i.i.d. process with $z_t \overset{D}{=} N(0,1)$, and z_t is independent of $\{x_{t-1}, x_{t-2}, \dots\}$ for all t. This type of process has been studied in detail by Ling (2004) and extended to a multivariate setting in Nielsen and Rahbek (2014) for the modelling interest rate dynamics; see also Hansen (2021).

V.4 Properties of univariate SREs

In this section, we consider properties of $\{X_t\}_{t\in\mathbb{Z}}$ given by the SRE in (V.1). With $V_t = (A_t, B_t)$, we note that the process $\{V_t\}_{t\in\mathbb{Z}}$ generates a natural filtration $\mathcal{F}_t = \sigma(V_t, V_{t-1}, \dots)$, and clearly, by independence, V_{t+1} is unpredictable with respect to $\mathcal{F}_t$. Likewise, we may consider cases where $X_t \in \mathcal{F}_t$ and X_t is independent of $\{V_{t+1}, V_{t+2}, \dots\}$. When the latter holds, we say that $\{X_t\}_{t\in\mathbb{Z}}$ is a *causal* process (with respect to $\mathcal{F}_t$). From the recursions in (**??**),

$$X_t = \sum_{i=0}^{t-1} \prod_{j=0}^{i-1} A_{t-j} B_{t-i} + \prod_{j=0}^{t-1} A_{t-j} X_0,$$

It shows up that the process $\{X_t\}_{t\in\mathbb{Z}}$ has a stationary causal solution when the last term term becomes asymptotically negligible, that is, the term vanishes as $t \to \infty$. In this case, X_t is given by the infinite series

$$X_t = \sum_{i=0}^{\infty} \prod_{j=0}^{i-1} A_{t-j} B_{t-i}, \tag{V.3}$$

which by construction is causal. Heuristically, in the one-dimensional case we may think of asymptotic negligibility as $\prod_{j=0}^{t-1} A_{t-j} \overset{P}{\to} 0$ as $t \to \infty$. Using Markov's inequality and that $\{A_t\}_{t\in\mathbb{Z}}$ is an i.i.d. process for any $\epsilon > 0$ and some $\delta > 0$,

$$\mathbb{P}\left(\left|\prod_{j=0}^{t-1} A_{t-j}\right| > \epsilon\right) \leq \frac{\mathbb{E}\left[\left|\prod_{j=0}^{t-1} A_{t-j}\right|^{\delta}\right]}{\epsilon^{\delta}} = \frac{(\mathbb{E}\left[|A_t|^{\delta}\right])^t}{\epsilon^{\delta}},$$

and we conclude that $\prod_{j=0}^{t-1} A_{t-j} \overset{P}{\to} 0$, if $\mathbb{E}[|A_t|^{\delta}] < 1$.

The next section provides conditions ensuring the existence of a stationary and ergodic causal solution to the SRE. For the ease of presentation we start out by considering results for the one dimensional case.

V.4.1 Stationarity and ergodicity

Let $\log^+(x) = \max\{\log(x), 0\}$, and note that, for instance, $\log^+(x) \leq |x|$. We have the following result.

Theorem V.4.1 (BDM, Theorem 2.1.3) *Consider the i.i.d. process* $\{(A_t, B_t)\}_{t\in\mathbb{Z}}$ *with* $(A_t, B_t)' \in \mathbb{R}^2$*-valued. Suppose that one of the following conditions holds:*

1. $\mathbb{P}(A_t = 0) > 0$;
2. $\mathbb{P}(A_t = 0) = 0$, $-\infty \leq \mathbb{E}[\log|A_t|] < 0$ *and* $\mathbb{E}[\log^+|B_t|] < \infty$.

Then the SRE in (V.1) has a stationary and ergodic causal solution given by (V.3).The infinite series in (V.3) converges almost surely for any $t \in \mathbb{Z}$.
On the contrary, assume that one of the following conditions holds:

1. $\mathbb{P}(A_t = 0) = 0$, $\mathbb{P}(B_t = 0) < 1$ *and* $0 \leq \mathbb{E}[\log|A_t|] < \infty$;
2. $\mathbb{E}[\log|A_t|] > -\infty$ *and* $\mathbb{E}[\log^+|B_t|] = \infty$.

Then no stationary causal solution to (V.1) exists.

One may loosely say that the condition $\mathbb{E}[\log|A_t|] < 0$ is the sufficient stationarity condition for the SRE. For $\delta > 0$, an application of Jensen's inequality gives that

$$\delta\mathbb{E}[\log|A_t|] = \mathbb{E}[\log|A_t|^\delta] \leq \log\mathbb{E}[|A_t|^\delta],$$

and we have that $\mathbb{E}[\log|A_t|] < 0$, provided that $\mathbb{E}[|A_t|^\delta] < 1$, which is the condition that we heuristically derived in the previous section.

Example V.4.1 (AR (1) ctd.) *We have that* $A_t = \rho$ *and* $\{B_t\}_{t\in\mathbb{Z}}$ *an i.i.d. process with* $\mathbb{E}[B_t] = 0$ *and* $\mathbb{E}[B_t^2] = \sigma^2 < \infty$*. Note that* $\mathbb{E}[\log|A_t|] = \log(|\rho|) < 0$*, if* $|\rho| < 1$*, and* $\mathbb{E}[\log^+|B_t|] \leq \mathbb{E}[|B_t|] \leq \sigma < \infty$*. Using Theorem V.4.1, we conclude that the AR(1) process is stationary and ergodic if* $|\rho| < 1$*. The causal solution is given by the infinite series*

$$x_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i},$$

identical to the one derived in (I.4) Likewise, we have that the process does not have a stationary causal solution, if $|\rho| \geq 1$.
Note that even if the AR(1) process has very heavy tailed errors ε_t*, it is still stationary and ergodic provided that* $|\rho| < 1$*, as we only require that*

$\mathbb{E}[\log^+ |\varepsilon_t|]$ *is finite. For instance, suppose that* ε_t *is Cauchy distributed such that* $\mathbb{E}[\varepsilon_t]$ *does not exist and* $\mathbb{E}[\varepsilon_t^2] = \infty$*. Then*

$$\begin{aligned}\mathbb{E}[\log^+ |\varepsilon_t|] &= \int_{-\infty}^{\infty} \log^+ |x| f_\varepsilon(x) dx \\ &= \int_{-\infty}^{\infty} \log^+ |x| \frac{1}{\pi(1+x^2)} dx \\ &= \frac{2}{\pi} \int_1^{\infty} \frac{\log(x)}{1+x^2} dx \approx 0.58 < \infty.\end{aligned}$$

Example V.4.2 (ARCH(1) ctd.) *If* $\alpha = 0$*, such that* $\mathbb{P}(A_t = 0) = 1$*, we have by Theorem V.4.1 that the ARCH(1) process is stationary and ergodic with* $X_t = B_t$ *(almost surely). Consider the case* $\alpha > 0$*. Let* $z \overset{D}{=} N(0,1)$*, and note that* $\mathbb{E}[\log^+ |B_t|] \leq \mathbb{E}[|B_t|] = \mathbb{E}[|\sqrt{\omega} z|] = \sqrt{\omega}\mathbb{E}[|z|] = \sqrt{\omega}\sqrt{2/\pi} < \infty$*. Moreover,* $\mathbb{E}[\log |A_t|] = \log(\alpha)/2 + \mathbb{E}[\log |z|]$*. We then have that* $\{x_t\}_{t\in\mathbb{Z}}$ *is stationary and ergodic if* $\log(\alpha)/2 + \mathbb{E}[\log |z|] < 0$*, or equivalently, if* $\alpha < \exp(-2\mathbb{E}[\log |z|]) = 3.56\ldots$*. In contrast, there exists no stationary causal solution to the SRE if* $\alpha \geq 3.56\ldots$*. In fact, in this case, if we consider (V.1) for only* $t \geq 1$ *with* x_0 *fixed, it can be shown that* $\sigma_t^2 = \omega + \alpha x_{t-1}^2 \to \infty$ *with probability tending to one as* $t \to \infty$*, and one may say that the conditional variance explodes; see Pedersen and Rahbek (2016, Supplementary Material) for more details.*
Alternatively, for the ARCH(1) process, we have that $x_t^2 = \sigma_t^2 z_t^2 = \omega z_t^2 + \alpha z_t^2 x_{t-1}^2$*. Hence, with* $X_t := x_t^2$*, we have the SRE in (V.1) with* $(A_t, B_t) = (\alpha z_t^2, \omega z_t^2)$*. Here* $\{X_t\}_{t\in\mathbb{Z}}$ *has a stationary and ergodic causal solution provided that* $E[\log |A_t|] = E[\log(\alpha z_t^2)] < 0$*. From (V.3) we have that*

$$x_t^2 = \sum_{i=0}^{\infty} \prod_{j=0}^{i-1} A_{t-j} B_{t-i} = \sum_{i=0}^{\infty} \prod_{j=0}^{i-1} (\alpha z_{t-j}^2) \left(\omega z_{t-i}^2\right) = \omega \sum_{i=0}^{\infty} \alpha^i \prod_{j=0}^{i} z_{t-i}^2,$$

identical to the expression for x_t^2 *found in (I.8).*

V.4.2 Moments and tail shape

Recall that the drift criterion considered in Part I provided sufficient conditions for finite moments of a particular order of the stationary solution to a Markov chain. Likewise, in terms of SREs we have the following result containing easy-to-verify conditions ensuring finite moments of the distribution of X_t.

Theorem V.4.2 (BDM, Lemmas 2.3.1-2.3.2) *Suppose that for* $d = 1$, *the SRE in* (*V.1*) *has a causal stationary solution. Moreover, assume that* $\mathbb{P}(A_t x + B_t = x) < 1$ *for all* $x \in \mathbb{R}$.

1. *If* $\mathbb{P}(A_t = 0) = \mathbb{P}(|A_t| = 1) = 0$ *and* $\mathbb{P}(B_t = 0) < 1$, *then for any* $s > 0$,
$$\mathbb{E}[e^{s|X_t|}] < \infty \text{ if and only if } \mathbb{P}(|A_t| < 1) = 1 \text{ and } \mathbb{E}[e^{s|B_t|}] < \infty.$$

2. *If* $\mathbb{P}(A_t = 0) = 0$ *and* $\mathbb{P}(B_t = 0) < 1$, *then*
$$\mathbb{E}[|X_t|^p] < \infty \text{ if and only if } \mathbb{E}[|A_t|^p] < 1 \text{ and } \mathbb{E}[|B_t|^p] < \infty.$$

Theorem V.4.2.2 provides conditions that are necessary and sufficient for X_t having a finite absolute moment of a given order. In contrast, the drift criterion in Part I only gave us sufficient conditions. For instance, in terms of and ARCH(1) process, in Example I.4.10, it was shown that $\mathbb{E}[X_t^2] < \infty$ if $\alpha < 1$ and $\mathbb{E}[X_t^4] < \infty$ if $\alpha < \sqrt{1/3}$. Below we show that these restrictions on α are sharp, in the sense that they are also necessary.

Example V.4.3 (AR(1) ctd.) *Recall that* $A_t = \rho$. *Let* $\rho \neq 0$ *with* $|\rho| < 1$ *(such that the AR(1) process is stationary and ergodic). Clearly for any fixed* $x \in \mathbb{R}$, $\mathbb{P}(A_t x + B_t = x) = \mathbb{P}(B_t = x(1-\rho)) < 1$, *since* $\mathbb{V}(B_t) > 0$. *Likewise,* $\mathbb{P}(B_t = 0) < 1$. *We conclude from Theorem V.4.2.1 that for any* $s > 0$, $E[e^{s|X_t|}] < \infty$ *if and only if* $\mathbb{E}[e^{s|B_t|}] < \infty$. *Likewise, from Theorem V.4.2.2,* $\mathbb{E}[|X_t|^p] < \infty$ *if and only if* $\mathbb{E}[|B_t|^p] < \infty$.

Example V.4.4 (ARCH(1) ctd.) *We focus on the stationary region* $0 < \alpha < 3.56\ldots$. *Since* (A_t, B_t) *are jointly normal, for any fixed* $x \in \mathbb{R}$, $A_t x + B_t \overset{D}{=} N(0, \alpha x^2 + \omega)$ *with* $\omega > 0$. *Consequently,* $\mathbb{P}(A_t x + B_t = x) < 1$ *for all* $x \in \mathbb{R}$. *Since* $A_t \overset{D}{=} N(0, \alpha)$ *with* $\alpha > 0$, $\mathbb{P}(|A_t| < 1) < 1$. *Hence, from Theorem V.4.2.1 there exists no* $s > 0$ *such that* $\mathbb{E}[e^{s|X_t|}] < \infty$. *On the other hand, by Gaussianity* $\mathbb{P}(A_t = 0) = \mathbb{P}(B_t = 0) = 0$ *and* $\mathbb{E}[|B_t|^p] < \infty$ *for any finite* p. *Consequently, by Theorem V.4.2.2,* $\mathbb{E}[|X_t|^p] < \infty$ *if and only if* $\mathbb{E}[|A_t|^p] < 1$. *For example,* $\mathbb{E}[X_t^2] < \infty$ *if and only if* $\alpha < 1$, *and* $\mathbb{E}[X_t^4] < \infty$ *if and only if* $\alpha < \sqrt{1/3}$.

The second part of Theorem V.4.2 suggests that X_t may be heavy-tailed in the sense that X_t may have infinite moments. We have the following result about the tail-shape of X_t, typically referred to as the Kesten-Goldie Theorem and refer to BDM (Section 2.4) for additional details (including the precise definitions of the constants) and arguments. In the following we use the notation that for positive functions f and g, $f(x) \sim g(x)$ as $x \to \infty$ if $\lim_{x\to\infty} f(x)/g(x) = 1$.

Theorem V.4.3 (BDM, Theorems 2.4.4 and 2.4.7) *Suppose that for* $d = 1$*, the SRE in* (*V.1*) *has a causal stationary solution. Moreover, assume that* $\mathbb{P}(A_t x + B_t = x) < 1$ *for all* $x \in \mathbb{R}$.

1. *If* $\mathbb{P}(A_t \geq 0) = 1$ *and the distribution of* $\log(A_t)$ *conditional on* $\{A_t > 0\}$ *is non-arithmetic*[4]*, and there exists a* $\kappa > 0$ *such that* $\mathbb{E}[A_t^\kappa] = 1$, $\mathbb{E}[|B_t|^\kappa] < \infty$ *and* $\mathbb{E}[A_t^\kappa \log_+(A_t)] < \infty$*, then there exist constants* $c_1, c_2 \geq 0$ *such that* $c_1 + c_2 > 0$ *and*

$$\mathbb{P}(X_t > x) \sim c_1 x^{-\kappa} \text{ and } \mathbb{P}(X_t < -x) \sim c_2 x^{-\kappa} \quad \text{as } x \to \infty. \tag{V.4}$$

2. *If* $\mathbb{P}(A_t < 0) > 0$ *and the distribution of* $\log |A_t|$ *conditional on* $\{A_t \neq 0\}$ *is non-arithmetic, and there exists a* $\kappa > 0$ *such that* $\mathbb{E}[|A_t|^\kappa] = 1$, $\mathbb{E}[|B_t|^\kappa] < \infty$ *and* $\mathbb{E}[|A_t|^\kappa \log_+ |A_t|] < \infty$*, then there exists a constant* $c_3 > 0$ *such that*

$$\mathbb{P}(X_t > x) \sim c_3 x^{-\kappa} \text{ and } \mathbb{P}(X_t < -x) \sim c_3 x^{-\kappa} \quad \text{as } x \to \infty. \tag{V.5}$$

Theorem V.4.3 describes the tail-shape of X_t. The result may be useful for quantifying the probabilities of extreme losses in actuarial sciences or risk management. The properties (V.4)-(V.5) mean that the distribution of X_t has power law, or so-called *regularly varying*, tails with $\kappa > 0$ denoting the *tail index*. Note that (V.4) or (V.5) imply that $\mathbb{E}[|X_t|^s] < \infty$ for $s < \kappa$, and $\mathbb{E}[|X_t|^s] = \infty$ for $s \geq \kappa$. This is important in terms of econometric theory, as finite moments are needed in order to apply the LLN (Theorem V.2.2) and CLT (V.2.3) when deriving limit results for estimators.

Example V.4.5 (AR(1) ctd.) *For the stationary case,* $|\rho| < 1$*, we note that there exists no* $\kappa > 0$ *such that* $\mathbb{E}[A_t^\kappa] = |\rho|^\kappa = 1$*. Hence, the results in Theorem V.4.3 do not apply to AR(1) processes. In fact, it can be shown that* X_t *can only have regularly varying tails if* B_t *has regularly varying tails.*

Example V.4.6 (ARCH(1) ctd.) *We focus again on the stationary region* $0 < \alpha < 3.56\ldots$*. We previously argued that* $\mathbb{P}(A_t x + B_t = x) < 1$ *for all* $x \in \mathbb{R}$*. Since* A_t *is Gaussian, it holds that* $\mathbb{P}(A_t < 0) > 0$*, and we hence seek to apply the second part of Theorem V.4.3. Clearly, since* $|A_t|$ *has a Lebesgue density, the distribution of* $\log |A_t|$ *conditional on* $\{A_t \neq 0\}$ *is non-arithmetic. Let* $A_t = \sqrt{\alpha} z_t$ *with* $z_t \overset{D}{=} N(0,1)$*, and note that for any*

[4] A distribution is said to be non-arithmetic if its support is not given by a set of the type $a\mathbb{Z}$, for some $a \geq 0$. Note that this condition is mild and holds, for instance, if A_t has a Lebesgue density.

$\kappa > 0$, $\mathbb{E}[|B_t|^\kappa] < \infty$ and $\mathbb{E}[|A_t|^\kappa \log_+ |A_t|] \leq \mathbb{E}[|A_t|^{1+\kappa}] < \infty$. Moreover, $\mathbb{E}[|A_t|^\kappa] = 1$ implies that $\alpha = (1/\mathbb{E}[|z_t|^\kappa])^{2/\kappa}$. Hence for such combinations of α and κ, X_t has regularly varying tails, that is, (V.5) holds. For instance, we have the pairs

$$(\kappa, \alpha) \in \{(1, \sqrt{\pi/2}), (2, 1), (4, 1/3^{1/2}), (6, 1/15^{1/3}), (8, 1/105^{1/4})\}.$$

In particular, the tails become thicker – that is, κ gets smaller – as α increases.

V.5 Properties of multivariate SREs

This section extends the previous results to the multivariate case. Inherently, the multivariate nature of the process makes the conditions more technical. Throughout, for a column vector $x \in \mathbb{R}^d$, let $|x|$ denote any norm. Moreover for any $d \times d$ matrix A, we consider the matrix norm induced by $|\cdot|$, $\|A\| = \sup_{x \in \mathbb{R}^d, |x|=1} |Ax|$ (the sup-norm).[5]

V.5.1 Stationarity and ergodicity

Recall from that for the univariate case in the previous section, we had (essentially) that the condition $\mathbb{E}[\log |A_t|] < 0$ implied stationarity of the SRE. In the multivariate setting ($d > 1$), the condition is different, reflecting that the matrix product $\prod_{i=1}^t A_i$ should vanish in order to ensure asymptotic negligibility. Here the relevant quantity of interest is the so-called top Lyapunov exponent associated with (V.1) as given by

$$\gamma = \inf_{t \geq 1} \frac{1}{t} \mathbb{E}[\log \| \prod_{i=1}^t A_i \|]. \tag{V.6}$$

We have the following result.

Theorem V.5.1 (BDM, Theorem 4.1.4) *Consider the i.i.d. process $\{(A_t, B_t)\}_{t \in \mathbb{Z}}$ with A_t a $d \times d$ real matrix and B_t $\mathbb{R}^d$-valued. Suppose that one of the following conditions hold:*

1. *$\mathbb{P}(A_t = 0) > 0$.*
2. *$\mathbb{E}[\log^+ \|A_t\|] < \infty$, $\mathbb{E}[\log^+ |B_t|] < \infty$ and the top Lyapunov exponent γ in (V.6) is strictly negative.*

[5] As an example, let $|x|$ denote the ℓ_2-norm, that is, $|x| = \sqrt{x'x}$, such that the induced matrix norm is the spectral norm, that is, the square-root of the largest eigenvalue of $A'A$.

Then there exists a stationary and ergodic causal solution to the SRE in (V.1). The solution is given by (V.3), and this infinite series converges almost surely.

Parallel to the univariate case, we may say that $\gamma < 0$ ensures stationarity of the SRE. Note that the matrix sup-norm is multiplicative, such that $\|\prod_{i=1}^{t} A_i\| \leq \prod_{i=1}^{t} \|A_i\|$. Hence we have that,

$$\frac{1}{t}\mathbb{E}[\log \|\prod_{i=1}^{t} A_i\|] \leq \frac{1}{t}\sum_{i=1}^{t} \mathbb{E}[\log \|A_i\|] = \mathbb{E}[\log \|A_t\|],$$

such that $\gamma < 0$ is implied by $\mathbb{E}[\log \|A_t\|] < 0$, which again (by Jensen's inequality) is implied by $\mathbb{E}[\|A_t\|] < 1$.

Example V.5.1 (VAR(1) ctd.) *Recall that $A_t = A$ is a constant matrix. If A is a zero matrix, then $X_t = B_t$ which clearly yields a stationary and ergodic process. We turn our attention to the non-trivial case where A is non-zero, and seek to apply the second part of Theorem V.5.1. Clearly $\mathbb{E}[\log^+ \|A_t\|] = \log^+ \|A\| < \infty$ and $\mathbb{E}[\log^+ |B_t|] \leq E[|B_t|] < \infty$, since B_t has a finite covariance matrix. It remains to find a condition such that the top Lyapunov exponent is strictly negative. Let $\rho(A)$ denote the spectral radius of A, and recall from Lemma A.3 in Part I that $\lim_{t\to\infty} \|A^t\|^{\frac{1}{t}} = \rho(A)$. We then have that $\gamma = \inf_{t\geq 1} \frac{1}{t}\mathbb{E}[\log \|\prod_{i=1}^{t} A_i\|] = \inf_{t\geq 1} \log \|A^t\|^{\frac{1}{t}} \leq \log \rho\,(A)$. Consequently $\gamma < 0$, if $\rho(A) < 1$. This condition is identical to the one derived in Section I.5, and we have the solution*

$$X_t = \sum_{i=0}^{\infty} A^i \varepsilon_{t-i}.$$

Furthermore, considering the VAR(k) process in Section I.5.1, identical arguments yield the stationarity condition $\rho(A) < 1$, with A denoting the companion matrix of the VAR(k) process.

Example V.5.2 (BEKK ctd.) *Recall that $A_t = m_t A$ with $m_t \stackrel{D}{=} N(0,1)$ and A a $(d \times d)$ matrix. Assume that A is non-zero. We then have that $P(A_t = 0) = 0$, and seek to apply the second part of Theorem V.5.1. Using that A_t and B_t have Gaussian entries, $\mathbb{E}[\log^+ \|A_t\|] < \infty$ and $\mathbb{E}[\log^+ |B_t|] \leq E[|B_t|] < \infty$, and it remains to find conditions ensuring that $\gamma < 0$. We have that $\mathbb{E}[\log \|\prod_{i=1}^{t} A_i\|] = \mathbb{E}[\log \|A^t \prod_{i=1}^{t} m_i\|] = \log \|A^t\| + t\mathbb{E}[\log(|m_t|)]$. Hence, $\gamma = \inf_{t\geq 1} \log \|A^t\|^{\frac{1}{t}} + \mathbb{E}[\log(|m_t|)] \leq \log \rho(A) + \mathbb{E}[\log(|m_t|)]$. We conclude that $\gamma < 0$ if $\rho(A) < \exp\left(-\mathbb{E}[\log(|m_t|)]\right) = \sqrt{3.56\ldots}$, where we recall the number $3.56\ldots$ from the stationarity condition in the univariate case in Example V.3.3.*

V.5.2 Moments and tail shape

Similar to the univariate case in Section V.4.2, we may consider conditions for finite moments of X_t as well as its tail shape. Inherently, this is a complicated task. In particular, the tail probabilities of a vector may be defined in multiple ways. For instance, each entry of X_t could have different tail index. This is for instance the case if A_t is a diagonal matrix such that X_t consists of d univariate SREs that may have different tail indexes. In such a case $\mathbb{E}[\|X_t\|^s] < \infty$ for any $s > 0$ smaller than the minimum of the tail indices. One way of defining tail probabilities of a vector is to consider the tail probability of linear combinations of X_t, and we refer BDM (Chapter 4) for many more details about so-called multivariate regular variation for multivariate SREs. For details about tail probabilities and finite moments of BEKK-ARCH processes, we refer to Matsui and Pedersen (2022). In terms of showing that $\|X_t\|$ has a a fininte moment of a given order, one may rely on the results in Tweedie (1988) which closely resembles the drift criterion but without requiring existence of a positive, continuous transition density.

V.6 Concluding remarks

The assumption that $\{(A_t, B_t)\}_{t\in\mathbb{Z}}$ being an i.i.d. process can be relaxed. In particular, in a univariate setting, Brandt (1986) considered conditions for stationarity and ergodicity for the case where $\{(A_t, B_t)\}_{t\in\mathbb{Z}}$ itself is stationary and ergodic. A multivariate extension of Brandt's result can be found in Bougerol and Picard (1992). The tail behavior of multivariate X_t and associated limit theory for partial sums of $\{X_t\}_{t\in\mathbb{Z}}$ is an active area of research. We refer to the recent textbook by Mikosch and Wintenberger (2024) for a comprehensive overview of results and examples.

References

Bougerol, P., & Picard, N., 1992, "Strict stationarity of generalized autoregressive processes", *The Annals of Statistics*, Vol. 20, pp. 1714–1730.

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

Buraczewski, D., Damek, E., & Mikosch, T., 2016, *Stochastic Models with Power-Law Tails: The Equation $X = AX + B$*, Springer.

Engle, R.F., 1982, "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, Vol. 50, pp. 987–1008.

Engle, R.F., & Kroner, K.F., 1995, "Multivariate simultaneous generalized ARCH", *Econometric Theory*, Vol. 11, pp. 122–150.

Francq, C. & Zakoïan, J.-M., 2019, *GARCH Models: Structure, Statistical Inference and Financial Applications*, 2nd edition. Wiley.

Hansen, A.L., 2021, "Modeling persistent interest rates with double-autoregressive processes", *Journal of Banking and Finance*, Vol. 133, 106302.

Ling, S., 2004, "Estimation and testing stationarity for double-autoregressive models", *Journal of the Royal Statistical Society: Series B*, Vol. 66, pp. 63–78.

Matsui, M., & Pedersen, R.S., 2022, "Characterization of the tail behavior of a class of BEKK processes: A stochastic recurrence equation approach", *Econometric Theory*, Vol. 38, pp. 1–34.

Mikosch, T., & Wintenberger, O., 2024, *Extreme Value Theory for Time Series: Models with Power-Law Tails*, Springer.

Nielsen, H.B., & Rahbek, A., 2014, "Unit root vector autoregression with volatility induced stationarity", *Journal of Empirical Finance*, Vol. 29, pp. 144–167.

Pedersen, R.S., & Rahbek, A., 2016, "Nonstationary GARCH with t-distributed innovations", *Economics Letters*, Vol. 138, pp. 19–21.

Tweedie, R.L., 1988, "Invariant measures for Markov chains with no irreducibility assumptions", *Journal of Applied Probability*, Vol. 25, pp. 275–285.