Anders Rahbek
Rasmus Søndergaard Pedersen
University of Copenhagen

August 2024

Part I
Introduction to Financial Time Series

In this chapter we introduce concepts needed for the probability analysis involved in the econometric analysis of financial time series as in Tsay (2010).

I.1 Time Series

Two of the most classic time series processes in financial econometrics are the autoregressive (AR) process, and the autoregressive conditional heteroskedastic (ARCH) process. These are briefly introduced in the next, before making concepts and ideas precise in the following sections.

I.1.1 AR process

The simplest AR process is of order one (AR(1)), with the AR(1) process given by

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

for $t = 1, 2, \ldots$ and with the recursion initiated in $x_0 = x \in \mathbb{R}$. The autoregressive parameter $\rho \in \mathbb{R}$, while the innovations ε_t are independently and identically distributed (i.i.d.), with a normal distribution with mean zero and variance $\sigma^2 > 0$, i.e. ε_t are i.i.d. $\mathrm{N}(0, \sigma^2)$. It follows that $\mathbb{E}[x_t | x_{t-1}] = \rho x_{t-1}$, while $\mathbb{V}(x_t | x_{t-1}) = \sigma^2$, and therefore the time-dependence, or dynamics, is modelled through the conditional mean of x_t given the past (observations).

Clearly the dynamic properties of the AR(1) process depend on the value of ρ. Note in this respect that simple recursion in (I.1) gives

$$x_t = \rho^t x + \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i}. \tag{I.2}$$

In particular, x_t is Gaussian distributed with unconditional mean $\mu_t = \rho^t x$ and variance

$$v_t = \left(1 + \rho^2 + \rho^4 + ... + \rho^{2(t-1)}\right)\sigma^2. \tag{I.3}$$

Both the mean and variance depends on t, and hence x_t's distribution is varying with t. At the same time, if $|\rho| < 1$, $\mu_t \to 0$, and using the well-known result for power series in Lemma I.3.1 below, we find

$$v_t = \frac{1-\rho^{2t}}{1-\rho^2}\sigma^2 \to \sigma^2 / \left(1-\rho^2\right).$$

Therefore, for $|\rho| < 1$, as $t \to \infty$, x_t will resemble the so-called *linear process*,

$$x_t^* = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}, \tag{I.4}$$

in terms of the sequence $\{\varepsilon_t\}_{t=...-1,0,1,2....}$ of i.i.d. $\mathrm{N}(0, \sigma^2)$ variables. The process $\{x_t^*\}_{t=0,1,2,...}$ is Gaussian distributed with $\mathbb{E}[x_t^*] = 0$ and $\mathbb{V}[x_t^*] = \sigma^2 / (1-\rho^2)$. The distribution of x_t^* does not depend on t, and is an example of a *stationary* process as detailed below. Thus if $|\rho| < 1$, x_t is *asymptotically stationary* in the sense that it resembles the stationary process x_t^* for large t. Note that x_t^* solves the recursion in (I.1) as can be seen by simple substitution. Also note that by giving x_0 the initial distribution as given by,

$$x_0 = \sum_{i=0}^{\infty} \rho^i \varepsilon_{0-i}, \tag{I.5}$$

x_t in (I.2) has the same distribution as the stationary solution x_t^*. That is, there exists a stationary solution to the AR(1) recursion, provided $|\rho| < 1$.

Remark I.1.1 *That the representation of x_t^* in (I.4) in terms of the infinite sum is well-defined follows by classic probability results, see e.g. Johansen (1996). While the actual form is appealing, it is not cruical to our theory; rather our focus is on showing that a stationary solution (here to the AR(1) recursion) exists, and that it has certain properties.*

I.1.2 ARCH process

The simplest ARCH process is the ARCH(1) which for $t = 1, 2, \ldots$ is given by

$$x_t = \sigma_t z_t \tag{I.6}$$
$$\sigma_t^2 = \omega + \alpha x_{t-1}^2 \tag{I.7}$$

with initial value $x_0 = x$ and where the innovations z_t are i.i.d. $\mathrm{N}(0,1)$. Moreover, conditionally on x_{t-1}, x_t is Gaussian distributed with mean zero and (conditional) variance σ_t^2. However, *unlike for the AR(1) process, this does not imply that x_t is Gaussian distributed unconditionally, or marginally.* Instead the marginal distribution of x_t is non-Gaussian, and – under regularity conditions discussed below – has in particular a more "fat tailed" distribution and can take "larger values" than expected if it was Gaussian distributed. The level parameter ω is strictly positive, $\omega > 0$, while the ARCH parameter $\alpha \geq 0$. Note that if $\alpha = 0$ then x_t is simply an i.i.d. $\mathrm{N}(0,\omega)$ sequence (conditionally and unconditionally). It should be emphasized that the conditional variance σ_t^2 is non-constant and stochastic.

The probabilistic behavior of the ARCH process x_t is complicated as for example the concept *stationarity* and *existence of moments* of x_t demands rather technical analysis. Using non-linear time series theory, we demonstrate below that while the ARCH sequence x_t is *uncorrelated,* it is *dependent.* Moreover, we will derive simple restrictions on the parameters (ω, α) for which x_t is well-behaved process in the sense that it is stationary, as well as having other desirable properties.

In line with the recursion for the AR(1), observe that the squared x_t satisfies a simple recursion,

$$\begin{aligned} x_t^2 &= \left(\omega + \alpha x_{t-1}^2\right) z_t^2 = \omega z_t + \alpha \left(\omega + \alpha x_{t-2}^2\right) z_t^2 z_{t-1}^2 \\ &= \omega \sum_{i=0}^{t-1} \alpha^i \prod_{j=0}^{i} z_{t-j}^2 + \omega \alpha^t x^2 \prod_{j=0}^{t-1} z_{t-j}^2. \end{aligned}$$

Hence, if $0 \leq \alpha < 1$, as $t \to \infty$, x_t^2 resembles (as $\alpha^t \to 0$, as $t \to \infty$) the stationary process,

$$(x_t^*)^2 = \omega \sum_{i=0}^{\infty} \alpha^i \prod_{j=0}^{i} z_{t-j}^2. \tag{I.8}$$

However, as we will demonstrate, while $\alpha < 1$ is sufficient for the existence of a stationary solution, it is not a necessary condition. That is, x_t indeed has a stationary solution for a range of values of $\alpha \geq 1$.

I.2 Conditional and unconditional moments

We introduce here some notation and concepts needed when discussing conditional and unconditional moments of the ARCH and AR processes. In particular, conditional expectations play a key role in these considerations and will be defined in terms of densities.

I.2.1 Expectations

Consider two random variables $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ for some $p, q \geq 1$, with well-defined densities $f(x)$ and $f(y)$, joint density $f(x, y)$ and, finally, well-defined conditional density $f(x|y) = f(x, y)/f(y)$. Recall that the expectation of X is given by

$$\mathbb{E}[X] = \int_{\mathbb{R}^p} x f(x)\, dx. \tag{I.9}$$

Likewise, the conditional expectation of X given $Y = y$, is given by

$$\mathbb{E}[X|Y = y] = \int_{\mathbb{R}^p} x f(x|y)\, dx, \tag{I.10}$$

which, by definition, is non-stochastic and depends on the value y. With $g(y) = \mathbb{E}[X|Y = y]$, we define furthermore the random variable, $\mathbb{E}[X|Y]$ as

$$E[X|Y] = g(Y). \tag{I.11}$$

Example I.2.1 *Consider $(X, Y)'$ bivariate $\mathrm{N}(\mu, \Omega)$ distributed, with mean μ and covariance matrix Ω given by,*

$$\mathbb{E}\left[\begin{pmatrix} X \\ Y \end{pmatrix}\right] = \mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \ \mathbb{V}\left[\begin{pmatrix} X \\ Y \end{pmatrix}\right] = \Omega = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix}.$$

In particular, X is $\mathrm{N}(\mu_x, \sigma_x^2)$ and Y is $\mathrm{N}(\mu_y, \sigma_y^2)$ distributed, while $\mathrm{Cov}[X, Y] = \sigma_{xy}$. Furthermore we have the important result that the conditional expectation of X given $Y = y$, is given by

$$\mathbb{E}[X|Y = y] = \mu_x + \omega(y - \mu_y) = \mu_{x|y}, \quad \textit{where } \omega = \sigma_{xy}/\sigma_y^2. \tag{I.12}$$

In fact, the conditional distribution of X given $Y = y$ is $\mathrm{N}\left(\mu_{x|y}, \sigma_{x|y}^2\right)$ distributed, with conditional variance,

$$\sigma_{x|y}^2 = \sigma_{xx}^2 - \omega\sigma_{yx}.$$

In particular, the conditional density is given by

$$f(x|y) = \frac{1}{\sqrt{2\pi\sigma_{x|y}^2}} \exp\left(-\frac{1}{2\sigma_{x|y}^2}\left(x - \mu_{x|y}\right)^2\right).$$

Moreover, $\mathbb{E}[X|Y] = \mu_x + \omega(Y - \mu_y)$ with

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[\mu_x + \omega(Y - \mu_y)] = \mu_x + \omega(\mathbb{E}[Y] - \mu_y) = \mu_x = \mathbb{E}[X]. \tag{I.13}$$

The fact that $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$ in the above example is a general feature of the conditional expectation, often referred to as the law of iterated expectations. We list here some well-known properties of conditional expectations which are simple to verify under the assumed setting of densities. A proof can be found in most probability theory books, see e.g. Durrett (2019, Ch.4.1.2).

Lemma I.2.1 *Consider the random variables X, Y and Z with joint density $f(x, y, z)$ and finite expectation. For the conditional expectation $\mathbb{E}[X|Y]$ the law of iterated expectations apply,*

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]. \tag{I.14}$$

If X and Y are independent,

$$\mathbb{E}[X|Y] = \mathbb{E}[X]. \tag{I.15}$$

Moreover,

$$\mathbb{E}[X|Y] = \mathbb{E}[(\mathbb{E}[X|Y, Z]|Y)], \quad \mathbb{E}[X|X] = X. \tag{I.16}$$

With g and h functions such that $g(Y)$ and $h(Y)$ take values in $\mathbb{R}$,

$$\mathbb{E}[(g(Y) + h(Y)X)|Y] = g(Y) + h(Y)\mathbb{E}[X|Y]. \tag{I.17}$$

Example I.2.2 *By definition of the AR process in (I.1)*

$$\mathbb{E}[x_t|x_{t-1}] = \mathbb{E}[(\rho x_{t-1} + \varepsilon_t)|x_{t-1}] = \rho\mathbb{E}[x_{t-1}|x_{t-1}] + \mathbb{E}[\varepsilon_t] = \rho x_{t-1},$$

where we have used (I.17), (I.16) and (I.15).

Example I.2.3 *By definition of the ARCH process in (I.6),*

$$\begin{aligned}
\mathbb{E}[x_t|x_{t-1}] &= \mathbb{E}\left[\left(\sqrt{[\omega + \alpha x_{t-1}^2]}z_t\right)|x_{t-1}\right] \\
&= \sqrt{[\omega + \alpha x_{t-1}^2]}\mathbb{E}[z_t|x_{t-1}] \\
&= \sqrt{[\sigma^2 + \alpha x_{t-1}^2]}\mathbb{E}[z_t] = 0.
\end{aligned}$$

Thus, provided $\mathbb{E}[|x_t|] < \infty$, the ARCH(1) has mean zero, $\mathbb{E}[x_t] = \mathbb{E}[\mathbb{E}[x_t|x_{t-1}]] = 0$.

Also note for later use when discussing for example the ARCH-implied so-called "Value-at-Risk", that the conditional distribution of X given a set $A = \{x : h(x) > a\}$, with $\mathbb{P}(X \in A) > 0$ and $h : \mathbb{R} \to \mathbb{R}$, has density

$$f(x|A) = \frac{f(x)}{P(X \in A)} \quad \text{for } h(x) > a,$$

such that,

$$\mathbb{E}[X|X \in A] = \int x f(x|A)\, dx.$$

Example I.2.4 *With x_t $\mathrm{N}(0, \sigma^2)$, distributed, $\mathbb{P}(x_t > 0) = \frac{1}{2}$, and density of the distribution of x_t conditional on $x_t > 0$,*

$$\begin{aligned} f(x_t|x_t > 0) &= \frac{f(x_t)}{P(x_t > 0)} \mathbb{I}(x_t > 0) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} x_t^2\right)}{\frac{1}{2}} \mathbb{I}(x_t > 0) \\ &= \sqrt{\frac{2}{\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} x_t^2\right) \mathbb{I}(x_t > 0) \end{aligned}$$

with $\mathbb{I}(\cdot)$ the indicator function. Next,

$$\begin{aligned} \mathbb{E}[x_t|x_t > 0] &= \int_{-\infty}^{\infty} x \sqrt{\frac{2}{\sigma^2\pi}} \exp\left(-\frac{1}{2\sigma^2} x^2\right) \mathbb{I}(x > 0)\, dx \\ &= 2 \int_0^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} x^2\right) dx \\ &= \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} x^2\right) dx \\ &= \mathbb{E}[|x_t|] = \sigma\sqrt{\frac{2}{\pi}}. \end{aligned}$$

Here we have used well-known properties of the Gaussian distribution, including symmetry, and $\mathbb{E}|Z| = \sqrt{2/\pi}$, with Z $\mathrm{N}(0, 1)$ distributed (and $x_t = \sigma Z$).

I.2.1.1 Some further comments on conditioning and moments

It is useful to discuss conditioning not only on x_{t-1} as in Example I.2.3 but also on all past variables $\mathcal{F}_{t-1} = (x_{t-1}, x_{t-2}..., x_0)$ (or, the σ-algebra, denoted $\sigma(x_{t-1}, x_{t-2}..., x_0)$). This is often referred to as 'conditioning on all past information' in the literature. For both the AR(1) and the ARCH(1) processes, the distribution of x_t conditional on past information up to time $t-1$ as given by the lagged variables in $\mathcal{F}_{t-1}$ depends only on x_{t-1}. This is commonly referred to as the *Markov property* and it plays an important role for the analysis of stochastic properties of $\{x_t\}$.

Note also at the same time that the simple specifications of the ARCH(1) and AR(1) are for empirical relevance extended in several possible ways, including adding more lags (x_{t-q}, $q \geq 2$) and non-linear functional forms in the conditional mean (AR) and variance (ARCH).

In terms of the definition of conditional expectations, we give meaning to $\mathbb{E}[x_t|\mathcal{F}_{t-1}]$ by setting $X = x_t$ and $Y = \mathcal{F}_{t-1}$, and often write it as $\mathbb{E}[x_t|\mathcal{F}_{t-1}]$,with $\mathcal{F}_{t-1} = (x_{t-1}, x_{t-2} \ldots, x_0)$. Thus as in the example for the ARCH(1) process x_t,

$$\mathbb{E}[x_t] = \mathbb{E}[\mathbb{E}[x_t|\mathcal{F}_{t-1}]] = 0.$$

That the calculations are identical, reflects the already mentioned fact that by the definition of the ARCH(1) process, the distribution of x_t conditional on $\mathcal{F}_{t-1} = (x_{t-1}, x_{t-2}, ..., x_0)$ depends only on x_{t-1}. Continuing with the ARCH(1) process , consider the correlation between x_t and x_{t-1},

$$\mathbb{E}[x_{t-1}x_t] = \mathbb{E}[\mathbb{E}[x_t x_{t-1}|\mathcal{F}_{t-1}]] = \mathbb{E}[x_{t-1}\mathbb{E}[x_t|\mathcal{F}_{t-1}]] = 0$$

and likewise,

$$\mathbb{E}[x_{t-k}x_t] = \mathbb{E}[\mathbb{E}[x_{t-k}x_t|\mathcal{F}_{t-k}]] = 0 \text{ for any } k \geq 1$$

Hence the ARCH(1) process is a mean zero and uncorrelated process. Note that this – unlike for uncorrelated Gaussian variables – *does not imply* that x_t and x_{t-1} are independent as e.g. σ_t^2, and hence x_t, depends on x_{t-1}^2.

Now turn to the second order moment, where, as $\mathbb{E}[x_t] = 0$,

$$\mathbb{V}[x_t] = \mathbb{E}\left[x_t^2\right] = \mathbb{E}\left[\mathbb{E}\left[x_t^2|\mathcal{F}_{t-1}\right]\right] = \mathbb{E}\left[\sigma_t^2\mathbb{E}\left[z_t^2\right]\right] = \mathbb{E}\left[\sigma_t^2\right] = \sigma^2 + \alpha\mathbb{E}\left[x_{t-1}^2\right] \tag{I.18}$$

Hence, if $\alpha < 1$ and $\mathbb{E}[x_t^2]$ is assumed constant such that $\mathbb{E}[x_t^2] = \mathbb{E}\left[x_{t-1}^2\right]$,

$$\mathbb{V}[x_t] = \mathbb{E}\left[x_t^2\right] = \frac{\sigma^2}{1-\alpha}. \tag{I.19}$$

Next, consider the "tail" behavior of x_t by considering 3. and 4. order moments. As odd moments of the $\mathrm{N}(0,1)$ distribution are zero it follows as above for the first order moment that,

$$\mathbb{E}\left[x_t^{2k+1}\right] = 0$$

i.e. all odd moments are zero. For the 4th order moment it follows that, if $\alpha < 1/\sqrt{3}$,

$$\mathbb{E}\left[x_t^4\right] = 3(\tfrac{1-\alpha^2}{1-3\alpha^2})\left(\mathbb{E}\left[x_t^2\right]\right)^2 > 3\left(\mathbb{E}(x_t^2)\right)^2. \tag{I.20}$$

As $\mathbb{E}[x_t^4] / (\mathbb{E}[x_t^2])^2 = 3$ for the Gaussian case, we conclude that the ARCH(1) process has *excess kurtosis* since $1 - \alpha^2 > 1 - 3\alpha^2$.

The above reflects that the existence of finite moments of a process x_t in particular for the case of non-linear time series depends on the parameter values. For the linear AR(1) process in (I.1), we find

$$\mathbb{E}[x_t] = \mathbb{E}[\mathbb{E}[x_t | \mathcal{F}_{t-1}]] = \rho \mathbb{E}[x_{t-1}],$$

hence if $\mathbb{E}[x_t]$ is constant, $\mathbb{E}[x_t] = 0$. Likewise,

$$\mathbb{V}[x_t] = \mathbb{E}\left[x_t^2\right] = \sigma^2 + \rho^2 \mathbb{E}\left[x_{t-1}^2\right] = \sigma^2 / \left(1 - \rho^2\right)$$

provided $\mathbb{E}[x_t^2]$ is constant and $|\rho| < 1$. In fact, all moments $\mathbb{E}\left[|x_t^k|\right] < \infty$, $k \geq 1$, with $|\rho| < 1$.

The "constant", or rather, non time-varying, moment assumption used repeatedly above is closely related to the concept of stationarity discussed next.

I.3 Stationarity and dependence

Above we considered moments $\mathbb{E}\left[x_t^k\right]$ for the ARCH(1) and AR(1) process assuming that they were identical for all time points. This leads to the concept of stationarity.

Definition I.3.1 *The process $\{X_t\}_{t=0,1,2,\ldots}$ is said to be stationary, or simply X_t is stationary, if for all t and h with $t, h \geq 0$, the joint distribution of $(X_t, \ldots, X_{t+h})$ does not depend on $t, t \geq 0$.*

Note that by definition for a stationary process with well-defined second order moments, the expectation $\mathbb{E}[X_t]$ and variance $\mathbb{V}[X_t]$ are constant, while the covariance between X_t and X_{t+h}, $\operatorname{Cov}(X_t, X_{t+h})$ depends only on h, and not on t.

Example I.3.1 *With x_t i.i.d. $\mathrm{N}(0, \sigma^2)$, then for $t, h \geq 0$,*

$$(x_t, \ldots, x_{t+h}) \text{ is } N_{h+1}(0, \Omega_h),$$

with $\Omega_h = \sigma^2 I_{h+1}$ where I_{h+1} is the $(h+1)$-dimensional identity matrix. This distribution does not depend on t and naturally the i.i.d. sequence is stationary.

This was a very simple example of a Gaussian process, where X_t is said to be Gaussian if $(X_t, ..., X_{t+h})$ is Gaussian distributed for all t and h. As the Gaussian distribution is characterized alone by the first two moments, it holds that X_t is stationary if, and only if, $\mathbb{E}[X_t]$ is constant and $\mathrm{Cov}[X_t, X_{t+h}] = v(h)$ that is, the covariance is a function of h and hence independent of t. Thus for Gaussian processes it is enough to consider the first two moments when discussing stationarity.

Example I.3.2 *The univariate Gaussian moving average process x_t of order 1, MA(1), is given by*

$$x_t = \varepsilon_t + \theta\varepsilon_{t-1},$$

with ε_t i.i.d. $\mathrm{N}(0, \sigma^2)$. In particular x_t is a stationary process with $\mathbb{E}[x_t] = 0$, $\mathbb{V}[x_t] = (1 + \theta^2)\sigma^2$, $\mathrm{Cov}(x_t, x_{t+1}) = \theta\sigma^2$ and

$$\mathrm{Cov}[x_t, x_{t+h}] = 0 \quad \text{for } h > 1.$$

Hence the MA(1) process is stationary as the Gaussian distribution is characterized fully by the first and second order moments.

In the next example we use that for power series:

Lemma I.3.1 *With $\phi \in \mathbb{R}$ and $\phi \neq 1$, then*

$$1 + \phi + \phi^2 + ... + \phi^n = \sum_{i=0}^{n} \phi^i = \left(1 - \phi^{n+1}\right) / (1 - \phi).$$

If moreover $|\phi| < 1$, $\phi^n \to 0$ as $n \to \infty$, and

$$\sum_{i=0}^{\infty} \phi^i = 1/(1 - \phi). \tag{I.21}$$

Example I.3.3 *Consider the solution x_t^* to the AR(1) process in (I.4). It follows that $x_t^* = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}, |\rho| < 1$, is Gaussian with $\mathbb{E}[x_t^*] = 0$, using (I.21) with $\phi = \rho^2 < 1$,*

$$\mathbb{V}[x_t^*] = \sum_{i=0}^{\infty} \rho^{2i}\sigma^2 = \sigma^2 / \left(1 - \rho^2\right).$$

Next, as $\mathrm{Cov}\left[\varepsilon_t, \varepsilon_{t+h}\right] = 0$ *for* $h \geq 1$*,*

$$\begin{aligned}\mathrm{Cov}\left[x_t^*, x_{t+h}^*\right] &= \mathbb{E}\left[x_t^* x_{t+h}^*\right] = \mathbb{E}\left[\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i} \sum_{i=0}^{\infty} \rho^i \varepsilon_{t+h-i}\right] \\ &= \mathbb{E}\left[\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i} \sum_{i=h}^{\infty} \rho^i \varepsilon_{t+h-i}\right] \\ &= \mathbb{E}\left[\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i} \sum_{j=0}^{\infty} \rho^i \varepsilon_{t-i} \rho^h\right] = \rho^h \mathbb{V}\left[x_t^*\right],\end{aligned}$$

Hence x_t^* *is indeed stationary. Note that all computations done here requires in particular* $\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}$ *to be well-defined. The process is an example of linear processes known from time series analysis and is well-defined for* $|\rho| < 1$ *and* ε_t *i.i.d.. The Markov chain theory below will show that it is in fact not needed to introduce the infinite sums and the explicit stationary solution* x_t^* *to discuss stationarity and dependence structure of the AR process.*

I.3.1 Dependence vanishing over time

The definition of stationarity addresses the joint distribution of the variables $X_{t+1}, \ldots, X_{t+h}$ for all h and t, but states nothing about dependence over time. If the stationary process $\{X_t\}_{t \in \mathbb{Z}}$ is dependent over time, with $\mathbb{E}\left[|X_t|^2\right] < \infty$, a key indicator often used for detection of dependence is the so-called autocorrelation. For a stationary process $X_t \in \mathbb{R}$, the so-called autocovariance function is given by

$$\mathrm{v}(h) = \mathrm{Cov}\left[X_t, X_{t+h}\right], \tag{I.22}$$

and the autocorrelation function (ACF) is defined by,

$$\mathrm{ACF}(h) = \mathrm{Corr}\left[X_t, X_{t+h}\right] = \frac{\mathrm{Cov}\left[X_t, X_{t+h}\right]}{\sqrt{\mathbb{V}\left[X_t\right] \mathbb{V}\left[X_{t+h}\right]}} = \frac{v(h)}{v(0)}, \tag{I.23}$$

where the last equality holds by stationarity. The functions $\rho\left(h\right)$ and $v\left(h\right)$ for various h describe the correlatedness, and hence indicates possible dependence over periods of time.

More generally, various kinds of dependence over time, often referred to as "mixing", or asymptotic independence, is used to describe vanishing dependence between X_t and X_{t+h} as h increases. This idea is crucial for time series and replaces the concept of independence. The idea is that a stationary process $(X_t)_{t=0,1,\ldots}$ is said to be mixing (or, ergodic), if for all t, h and sets

A, B,

$$P((X_0,\dots,X_t)\in A,(X_h,\dots,X_{t+h})\in B)\to \qquad \text{(I.24)}$$
$$P((X_0,\dots,X_t)\in A)P((X_0,\dots,X_t)\in B) \quad h\to\infty.$$

Here "mixing" is used loosely and the literature includes many different forms of "mixing" (including α and β-mixing) definitions; but they all share the idea that dependence between X_t and X_{t+h} vanishes as h increases. Or stated differently, events removed far in time from one another are independent. Moreover, they imply various types of law of large numbers (LLNs) apply.

Below we discuss the so-called drift criterion from Markov chain theory which provides a useful tool to establish conditions under which LLNs and central limit theorems (CLTs) hold for time series.

Before turning to the drift criterion, we note the following result which relates mixing and correlations

Lemma I.3.2 *If the stationary process $\{X_t\}_{t=0,1,2,..}$ satisfies (I.24) and has finite variance, then the covariance $v(h)=\mathrm{Cov}\,[X_t,X_{t+h}]$ tends to zero as $h\to\infty$. On the other hand, if $\{X_t\}$ is a stationary Gaussian process for which $v(h)\to 0$ as $h\to\infty$, then X_t satisfies (I.24).*

Example I.3.4 *It follows by Example I.3.2 that the MA(1) process satisfies (I.24), and likewise, if $|\rho|<1$, the AR(1) process x_t^* does.*

I.4 Drift criterion from Markov chain theory

If a Markov chain $\{X_t\}_{t=0,1,2,...}$ satisfies the drift criterion, a first important implication is that the initial value, X_0, can be given a distribution such that X_t is stationary. This resembles the considerations made for the AR(1) process x_t where as shown one can explicitly choose an initial distribution of x_0 such that $x_t \overset{D}{=} x_t^*$ and hence stationary. Second, the drift criterion implies finiteness of certain moments for the stationary version. Moreover, variants of the law of large numbers (LLN) and the central limit theorem (CLT) apply.

In short, the drift criterion is very helpful in many ways and a powerful tool for time series analysis. The introduction here is based on Meyn and Tweedie (1993) and Tjøstheim (1990).

I.4.0.1 Assumptions

A common key feature of the AR(1) and ARCH(1) processes is that with X_t denoting either of the two, then the distribution of,

$$X_t \text{ conditional on } (X_{t-1}, X_{t-2}, ..., X_0)$$

depends only on X_{t-1}. More precisely, in the AR(1) case x_t conditionally on x_{t-1}, is $\mathrm{N}(\rho x_{t-1}, \sigma^2)$ distributed, while for the ARCH(1) x_t conditionally on x_{t-1}, is $\mathrm{N}(0, \sigma_t^2)$ distributed with $\sigma_t^2 = \sigma^2 + \alpha x_{t-1}^2$. In both cases, the conditional distribution has a Gaussian density which has some attractive features.

We make the following assumption:

Assumption I.4.1 *Assume that for $(X_t)_{t=0,1,2,...}$ with $X_t \in \mathbb{R}^p$ it holds that:*

(i) *the conditional distribution of X_t given $(X_{t-1}, X_{t-2}, ..., X_0)$ depends only on X_{t-1}, that is*

$$X_t|X_{t-1}, X_{t-2}, ..., X_0 \overset{D}{=} X_t|X_{t-1}.$$

(ii) *the conditional distribution of X_t given X_{t-n}, for some $n \geq 1$, has a positive (n−step) conditional density $f(y|x) > 0$, which is continuous in both arguments.*

Note that (i) implies that $\{X_t\}_{t=0,1,2,...}$ is a Markov chain on $\mathbb{R}^p$, sometimes referred to as a Markov chain on a general state space. Also note that the condition (ii) of continuity is often simple to validate, in particular for $k = 1$.

Example I.4.1 *For the AR(1) process in (I.1), x_t conditional on x_{t-1} has density*

$$f(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\tfrac{(y-\rho)^2}{2\sigma^2}\right),$$

which is positive and continuous in y and x. Often we simply write the conditional density in terms of x_t,

$$f(x_t|x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\tfrac{(x_t-\rho x_{t-1})^2}{2\sigma^2}\right).$$

Example I.4.2 *For the ARCH process in (I.6), x_t conditional on x_{t-1} has the Gaussian density,*

$$f(x_t|x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\tfrac{1}{2\sigma_t^2} x_t^2\right), \quad \sigma_t^2 = \sigma^2 + \alpha x_{t-1}^2$$

which is positive and continuous.

For the next example recall initially the following well-known result from probability analysis:

Lemma I.4.1 *If a real variable X, $X \in \mathbb{R}$, has density $f(x)$, then $Y = cX$, with $c \neq 0$ a constant, has density $\frac{1}{\sqrt{c^2}} f\left(\frac{y}{c}\right)$. Moreover, if $\mathbb{V}[X] < \infty$, then $\mathbb{E}[Y] = c\mathbb{E}[X]$ and $\mathbb{V}[Y] = c^2 \mathbb{V}[X]$.*

Example I.4.3 *In the ARCH process x_t in (I.6) the assumption of z_t i.i.d.$N(0,1)$ is sometimes replaced by the assumption that z_t is i.i.d. with $\mathbb{E}[z_t] = 0$, $\mathbb{V}[z_t] = 1$ and t_v-distribution scaled by $\sqrt{\frac{v-2}{v}}$. Here $v > 2$ and denotes the degrees of freedom. An ARCH process defined this way satisfies Assumption I.4.1.*

To see this note first that if X is t_v-distributed with $v > 2$, then X has $\mathbb{E}[X] = 0$ and $\mathbb{V}[X] = v/(v-2)$. Moreover, X has density,

$$f(x) = \frac{\gamma(v)}{\sqrt{v\pi}} \left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)},$$

where the constant $\gamma(v) = \Gamma\left(\frac{v+1}{2}\right) / \Gamma\left(\frac{v}{2}\right)$, with $\Gamma(\cdot)$ the so-called Gamma function. As $\mathbb{V}[X] = v/(v-2)$, then using Lemma I.4.1, $z_t = \left(\sqrt{\frac{v-2}{v}}\right) X$ satisfies $\mathbb{V}[z_t] = 1$ and $\mathbb{E}(z_t) = 0$. By simple insertion z_t has density

$$f(z) = \frac{\gamma(v)}{\sqrt{(v-2)\pi}} \left(1 + \frac{z^2}{(v-2)}\right)^{-\left(\frac{v+1}{2}\right)}.$$

Next, by the ARCH equation $x_t = \sigma_t z_t$, and using Lemma I.4.1 again, x_t conditional on x_{t-1} has density,

$$f(x_t | x_{t-1}) = \frac{\gamma(v)}{\sqrt{\sigma_t^2 (v-2)\pi}} \left(1 + \frac{x_t^2}{\sigma_t^2 (v-2)}\right)^{-\left(\frac{v+1}{2}\right)}, \quad \sigma_t^2 = \sigma^2 + \alpha x_{t-1}^2.$$

I.4.0.2 Drift function

Next, we define *a drift function* for a process X_t satisfying Assumption I.4.1. With X_t a time series, a drift function for X_t is some function $\delta(X_t)$, where $\delta(X_t) \geq 1$ and which is not identically ∞. The choice of drift function is quite flexible, but a key example is the next.

Example I.4.4 *A much used drift function in the analysis of univariate AR and ARCH processes with $X_t \in \mathbb{R}$, is*

$$\delta(X_t) = 1 + X_t^2,$$

while, if $X_t = (X_{1t}, ..., X_{pt})' \in \mathbb{R}^p$,

$$\delta(X_t) = 1 + X_t' X_t = 1 + \sum_{i=1}^{p} X_{it}^2.$$

The role of such a drift function is to measure the dynamics, or the *drift* of X_t, by studying the dynamics of $\delta(X_t)$ instead of X_t itself. considering the conditional expectation of $\delta(X_t)$ given X_{t-1} or some more "distant" past value of X_t, say X_{t-m}. That is we are interested in studying the dynamics of

$$\mathbb{E}\left(\delta(X_t) \,|X_{t-m}\right),$$

for some m, where typically $m = 1$ is used.

Example I.4.5 *For the AR(1) process* x_t *with* $\delta(x_t) = 1 + x_t^2$,

$$\begin{aligned}
\mathbb{E}\left[\delta(x_t)\,|x_{t-1}\right] &= \mathbb{E}\left[1 + (\rho x_{t-1} + \varepsilon_t)^2\,|x_{t-1}\right] \\
&= 1 + \rho^2 \mathbb{E}\left[x_{t-1}^2 | x_{t-1}\right] + 2\rho x_{t-1} \mathbb{E}\left[\varepsilon_t | x_{t-1}\right] + \mathbb{E}\left[\varepsilon_t^2 | x_{t-1}\right] \\
&= 1 + \rho^2 x_{t-1}^2 + 2\rho x_{t-1} \mathbb{E}\left[\varepsilon_t\right] + \mathbb{E}\left[\varepsilon_t^2\right] \\
&= 1 + \sigma^2 + \rho^2 x_{t-1}^2 \\
&= \rho^2 \delta(x_{t-1}) + c, \qquad \text{(I.25)}
\end{aligned}$$

where the constant c *is given by* $c = (1 - \rho^2 + \sigma^2)$. *Thus, apart from the constant* c, *we obtain what mimics a simple first order autoregression in* $\delta(x_t)$. *That is, we may write*

$$\delta(x_t) = \rho^2 \delta(x_{t-1}) + c + \eta_t,$$

with $\eta_t = (\delta(x_t) - \mathbb{E}(\delta(x_t)\,|x_{t-1}))$, *such that by definition* $\mathbb{E}[\eta_t] = 0$. *Thus if* $\rho^2 < 1$, $\delta(x_t)$ *resembles a stationary AR(1) process.*

Example I.4.6 *With ARCH(1) process in (I.6) and* $\delta(x_t) = 1 + x_t^2$,

$$\begin{aligned}
\mathbb{E}\left(\delta(x_t)\,|x_{t-1}\right) &= \mathbb{E}\left[1 + \sigma_t^2 z_t^2 | x_{t-1}\right] \\
&= 1 + \left(\sigma^2 + \alpha x_{t-1}^2\right) \mathbb{E}\left[z_t^2 | x_{t-1}\right] \\
&= 1 + \alpha x_{t-1}^2 + \sigma^2 \\
&= \alpha \delta(x_{t-1}) + c,
\end{aligned}$$

where $c = (1 + \sigma^2 - \alpha)$. *Thus as before in the AR example, we can interpretate this as a simple autoregression in* $\delta(x_{t-1})$ *with autoregressive coefficient* α.

For the dynamics of the drift function we make the following assumption:

Assumption I.4.2 *Assume that $\{X_t\}_{t=0,1,2,..}$, with $X_t \in \mathbb{R}^p$, satisfies Assumption I.4.1. With drift function δ, $\delta(X_t) \geq 1$, assume that there are positive constants M, C and ϕ, $\phi < 1$, such that for some $m \geq 1$,*

$$\begin{array}{ll} (i) & \mathbb{E}[\delta(X_{t+m})|X_t = X] \leq \phi\delta(X) \quad \text{for } X'X > M, \\ (ii) & \mathbb{E}[\delta(X_{t+m})|X_t = X] \leq C < \infty \quad \text{for } X'X \leq M. \end{array}$$

If $\{X_t\}$ satisfies Assumption I.4.2 then we say that X_t satisfies the drift criterion with drift function $\delta(\cdot)$. Note also that a simple way to verify *(i)*, is to show that if $\mathbb{E}[\delta(X_{t+m})|X_t = X] \leq g(X)$, say, then *(i)* holds if

$$g(X)/\delta(X) \to \phi \text{ as } X'X \to \infty.$$

Example I.4.7 *The AR process x_t in satisfies the drift criterion if $\rho^2 < 1$ with $\delta(x_t) = 1 + x_t^2$ with x_{t-1}^2 chosen large. Recall from Example I.4.4,*

$$\mathbb{E}[\delta(x_t)|x_{t-1}] = \rho^2\delta(x_{t-1}) + c, \quad \text{with } c = 1 - \rho^2 + \sigma^2.$$

hence (i) holds with $\rho^2 < \phi < 1$ for $x_{t-1}^2 > M$, M large enough. For $x_{t-1}^2 \leq M$,

$$\mathbb{E}(\delta(x_t)|x_{t-1}) = \rho^2\delta(x_{t-1}) + c \leq \rho^2\delta(M) + c = C.$$

Example I.4.8 *The ARCH process x_t satisfies the drift criterion if $\alpha < 1$ with $\delta(x_{t-1}) = 1 + x_{t-1}^2$. By Example I.4.6,*

$$\mathbb{E}(\delta(x_t)|x_{t-1}) = \alpha\delta(x_{t-1}) + c, \quad \text{with } c = \left(1 - \alpha + \sigma^2\right),$$

and we see that the considerations in Example I.4.7 can be applied here with $\alpha = \rho^2$.

We are now in position to state a powerful result from Tjøstheim (1990) and Jensen and Rahbek (2007):

Theorem I.4.1 *Assume that $\{X_t\}_{t\geq 0}$ satisfies Assumption I.4.2 with drift function δ. Then X_0 can be given an initial distribution such that X_t initiated in X_0 is stationary. With X_t^* denoting the stationary version, $\mathbb{E}[\delta(X_t^*)] < \infty$. Moreover, X_t is mixing in the sense that, for any initial value X_0, the LLN in Theorem I.4.2 below holds.* Restate

More precisely, X_t satisfying Theorem I.4.1 is referred to as being "geometrically ergodic" in the Markov chain literature, see e.g. Tjøstheim (1990) and Francq and Zakoian (2019) for details. The idea is that, with $f^{(n)}(y|x)$ denoting the (n-step) density of X_{t+n} conditional on X_t, and $f^*(\cdot)$ the density of the stationary solution X_t^*, then

$$f^{(n)}(y|x) \to f^*(y)$$

exponentially fast as $n \to \infty$. And, importantly, the exponential speed implies that a LLN (as well as a CLT) applies.

Thus, if X_t satisfies the drift criterion, not only can the process be considered stationary (by giving X_0 the correct distribution) and geometrically ergodic, but also the LLN applies, independently of the initial value. Moreover, as $\mathbb{E}(\delta(X_t^*)) < \infty$, any moments of X_t which are bounded by the drift function δ are finite.

Example I.4.9 *For the AR(1) process we may conclude from Example I.4.5 that $\mathbb{E}[x_t^{*2}] < \infty$, and that the law of large numbers apply to x_t by Theorem I.4.1. While this illustrates the results, this conclusion is not surprising as we already know that with $\rho^2 < 1$, then x_t has a stationary representation x_t^*, and since it is Gaussian, in fact, $\mathbb{E}\left[(x_t^*)^{2k}\right] < \infty$ for any k. To give an understanding of the role of initial value x_0, use that simple recursion gives*

$$x_{t+n} = \rho^n x_t + \sum_{i=0}^{n-1} \rho^i \varepsilon_{t+n-i},$$

Hence the n-step transition density is given by,

$$f^{(n)}(y|x) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(y-\rho^n x)^2}{2\sigma_n^2}\right), \qquad \sigma_n^2 = \sigma^2 \frac{1-\rho^{2n}}{1-\rho^2}.$$

Clearly, $f^{(n)}(y|x) \to f^*(y)$, corresponding to the stationary version,

$$x_t^* = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i} \overset{D}{=} \mathrm{N}\left(0, \sigma^2 \frac{1}{1-\rho^2}\right).$$

Observe in particular that $\rho^m \to 0$ exponentially fast.

Example I.4.10 *For the ARCH(1) process,*

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

with z_t i.i.d.$N(0,1)$ we conclude that if $0 \leq \alpha < 1$ then x_t has a stationary solution with $\mathbb{E}[x_t^{*2}] < \infty$. Hence any moments of order lower than 2 are finite, for example $\mathbb{E}[|x_t^*|] < \infty$ since $|x| \leq \delta(x) = 1 + x^2$. However, we do not know if for example x_t has finite fourth order moments, $\mathbb{E}[x_t^4] < \infty$. To find out under which conditions this holds we need to consider a drift function from which we can conclude this. An example is $\delta(x_t) = 1 + x_t^4$, where using $\mathbb{E}[z_t^4] = 3$, we find,

$$\begin{aligned}
\mathbb{E}[\delta(x_t)|x_{t-1}] &= 1 + \left(\sigma^2 + \alpha x_{t-1}^2\right)^2 \mathbb{E}\left[z_t^4\right] \\
&= 1 + 3\left(\sigma^4 + 2\alpha\sigma^2 x_{t-1}^2 + \alpha^2 x_{t-1}^4\right) \\
&= 3\alpha^2\left(1 + x_{t-1}^4\right) + \left(1 - 3\alpha^2 + 3\sigma^4\right) + 6\alpha\sigma^2 x_{t-1}^2 \\
&= 3\alpha^2 \delta(x_{t-1}) + c\left(x_{t-1}^2\right),
\end{aligned}$$

where $c\left(x_{t-1}^2\right) = c + 6\alpha\sigma^2 x_{t-1}^2$, with $c = \left(1 - 3\alpha^2 + 3\sigma^4\right)$. We thus need to choose α so small that $3\alpha^2 < 1$. Hence the conclusion is that while a stationary x_t exists for $\alpha < 1$ and $\mathbb{E}[x_t^{*2}] < \infty$ in this case, we need to restrict α further to have fourth order moments. More precisely, and as already indicated, provided $\alpha < 1/\sqrt{3} \simeq 0.56$ then $\mathbb{E}\left[(x_t^*)^4\right] < \infty$.

The considerations for the ARCH(1) illustrated that the value of α in the conditional variance $\sigma_t^2 = \sigma^2 + \alpha x_{t-1}^2$ was crucial for stationarity of x_t and also for the existence of finite moments of x_t. This is a typical feature of non-linear time series where parameter values have implications for interpretation in terms of both stationarity and finite moments.

A key implication of Theorem I.4.1 is that X_t is geometrically ergodic and the following LLN applies from Jensen and Rahbek (2007).

Theorem I.4.2 *Assume that with $X_t \in \mathbb{R}^p$, $\{X_t\}_{t=0,1,2,3,...}$ is a geometrically ergodic Markov chain with stationary solution $\{X_t^*\}$. Assume furthermore that the function $g : \mathbb{R}^{p(m+1)} \to \mathbb{R}$, $m \geq 0$, satisfies $\mathbb{E}\left[\left|g\left(X_t^*, X_{t-1}^*, ..., X_{t-m}^*\right)\right|\right] < \infty$, then as $T \to \infty$,*

$$\frac{1}{T}\sum_{t=1}^{T} g\left(X_t, X_{t-1}, ..., X_{t-m}\right) \xrightarrow{P} \mathbb{E}\left[g\left(X_t^*, X_{t-1}^*, ..., X_{t-m}^*\right)\right]. \tag{I.26}$$

We emphasize that the LLN holds for any initial value X_0, which is important for the later statistical analysis.

Note in that respect that $\mathbb{E}[\cdot]$ in (I.26) means the expectation under the stationary measure (X_t^* stationary), while in the average

$$\frac{1}{T}\sum_{t=1}^{T} g\left(X_t, X_{t-1}, ..., X_{t-m}\right),$$

X_0 does not have to be equipped with the stationary distribution and X_t therefore not stationary. Henceforth, unless important to make the distinction between X_t and X_t^*, we may sometimes just write

$$\frac{1}{T}\sum_{t=1}^{T} g\left(X_t, X_{t-1}, ..., X_{t-m}\right) \xrightarrow{P} \mathbb{E}\left[g\left(X_t, X_{t-1}, ..., X_{t-m}\right)\right].$$

Example I.4.11 *With X_t is univariate, and geometrically ergodic with finite second order moments, then*

$$\frac{1}{T}\sum_{t=1}^{T} X_t^2 \xrightarrow{P} \mathbb{E}\left[X_t^2\right] \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T} X_t X_{t-1} \xrightarrow{P} \mathbb{E}\left[X_t X_{t-1}\right],$$

by applying Theorem I.4.2 with

$$g\left(X_t\right) = X_t^2 \text{ and } g\left(X_t, X_{t-1}\right) = X_t X_{t-1}.$$

Example I.4.12 *If $\mathbb{E}\left[X_t\right] = 0$, such that $\mathbb{V}\left[X_t\right] = \mathbb{E}\left[X_t^2\right]$ and $\mathrm{Cov}\left[X_t, X_{t+h}\right] = \mathbb{E}\left[X_t X_{t+h}\right]$, it follows that the empirical autocorrelation function, see (I.23), as defined by,*

$$\widehat{\rho}(h) \equiv \frac{\frac{1}{T}\sum_{t=1}^{T-h} X_t X_{t+h}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} X_t^2 \frac{1}{T}\sum_{t=1}^{T-h} X_{t+h}^2}}, \tag{I.27}$$

will converge in probability to (the theoretical) $\rho\left(h\right)$ by Theorem I.4.2 motivating that most software for time series allows one to compute these directly.

As emphasized for the AR(1) process x_t, the autoregressive coefficient ρ is the key parameter to an understanding of the dynamics. We know that if $|\rho| < 1$, x_t is stationary (or, a stationary solution exists) and with ε_t Gaussian all moments are finite. At the same time we know that the restriction of $|\rho| < 1$ is crucial in the sense that if $\rho = 1$, as often found empirically, then x_t is a so-called "unit-root" process which is non-stationary. Surprisingly this is not so for the ARCH(1) process, where $\alpha = 1$ still allows x_t to be stationary. However, with $\alpha = 1$, then only moments up to order one are finite, $\mathbb{E}\left[|x_t|\right] < \infty$. That is, $\alpha = 1$ implies stationarity but for a process with infinite variance.

More precisely we can make the following table where we can divide the interval for α as follows:

ARCH process x_t defined in (I.6):	
$x_t = \sigma_t z_t,\ \ \sigma_t = \sigma^2 + \alpha x_{t-1}^2$ and z_t $i.i.d.\mathrm{N}(0,1)$.	
Stationary for $0 \leq \alpha < 3.56$	
Finite moments:	
$\frac{\pi}{2} \leq \alpha < 3.56$	$\mathbb{E}\left[\lvert x_t\rvert^{\delta}\right] < \infty$, some $\delta \in (0,1)$
$1 \leq \alpha < \frac{\pi}{2}$	$\mathbb{E}\left[\lvert x_t\rvert\right] < \infty$
$\frac{1}{\sqrt{3}} \leq \alpha < 1$	$\mathbb{E}\left[x_t^2\right] < \infty$
$0 \leq \alpha < \frac{1}{\sqrt{3}}$	$\mathbb{E}\left[x_t^4\right] < \infty$

Actually the number 3.56 is an approximation of $\exp(-\mathbb{E}\left[\log\left(z^2\right)\right])$, with z $\mathrm{N}(0,1)$ distributed. In fact, if z_t has another distribution such as the t_v distribution mentioned in Example I.4.3 the intervals above would change.

I.4.1 Central limit theorem

As noted, a powerful implication of Theorem I.4.1 for X_t is that independently of the initial value, X_0 the LLN in Theorem I.4.2 applies. The following theorem generalizes this to also hold for the central limit theorem (CLT) in Meyn and Tweedie (1993, ch. 17). More precisely, we have for Y_t a continuous function of X_t and, possibly, its lagged values, that is, $Y_t = f\left(X_t, X_{t-1}, ..., X_{t-m}\right)$:

Theorem I.4.3 (Meyn and Tweedie, 1993, Theorem 17.0.1) *Assume that Theorem I.4.1 applies to* $\{X_t\}_{t\geq 0}$, $Y_t = f\left(X_t, X_{t-1}, ..., X_{t-m}\right)$. *With* $Y_t^* = f\left(X_t^*, X_{t-1}^*, ..., X_{t-m}^*\right)$, *suppose that* $\mathbb{E}\left[Y_t^*\right] = 0$, $\mathbb{E}\left[\left(Y_t^*\right)^2\right] < \infty$ *and*

$$\gamma = \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\left(\sum_{t=1}^{T} Y_t^*\right)^2\right] > 0.$$

Then as $T \to \infty$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_t \xrightarrow{D} N(0,\gamma).$$

A different version of the CLT is a corrollary from Brown (1971) and we will often apply this (or a variant thereof) when discussing asymptotic normality later in the statistical analysis. As above, consider the mapping $Y_t = f\left(X_t, X_{t-1}, ..., X_{t-m}\right)$ with $f(\cdot)$ continuous, then Y_t is an example of a Martingale difference (MGD) sequence wrt. $\mathcal{F}_t = (X_t, ..., X_0)$, if $\mathbb{E}\left[Y_t | \mathcal{F}_{t-1}\right] = 0$.

Theorem I.4.4 (Corollary to Brown, 1971) *For a given sequence* $\{X_t\}_{t\geq 0}$*, consider* $Y_t = f(X_t, X_{t-1}, ..., X_{t-m})$*,* $f(\cdot)$ *continuous, with* $\mathbb{E}[Y_t|\mathcal{F}_{t-1}] = 0$*, where* $\mathcal{F}_t = (X_t, ..., X_0)$*. If, as* $T \to \infty$*,*

$$\text{(i): } T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[Y_t^2|\mathcal{F}_{t-1}\right] \xrightarrow{P} \sigma^2 > 0$$

and either (ii) or (ii') hold,

$$\text{(ii): } T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[Y_t^2\mathbb{I}\left(|Y_t| > \delta T^{1/2}\right)\right] \to 0,$$

$$\text{(ii'): } T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[Y_t^2\mathbb{I}\left(|Y_t| > \delta T^{1/2}\right)|\mathcal{F}_{t-1}\right] \xrightarrow{P} 0,$$

for any $\delta > 0$*, then* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_t \xrightarrow{D} N(0, \sigma^2)$*.*

Note that if $\{X_t\}_{t\geq 0}$ satisfies Theorem I.4.1, and $\mathbb{E}[Y_t^{*2}] < \infty$, then (i) in Theorem I.4.4 holds by the LLN in Theorem I.4.2. As to the so-called "Lindeberg"-type conditions in (ii) and (ii'), note that even though (ii) is "classic", also (ii') is stated here in as it often is useful in the context of time series. Also one may note that (ii) holds under stationarity (and dominated convergence), since if Y_t is stationary with $\mathbb{E}[|Y_t|^{2+\eta}] < \infty$, for some $\eta > 0$, we have

$$T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[Y_t^2\mathbb{I}\left(|Y_t| > \delta T^{1/2}\right)\right] \leq \frac{1}{\left(\delta T^{1/2}\right)^{\eta}}\mathbb{E}\left[|Y_t|^{2+\eta}\right] \to 0.$$

Various different versions of the CLT for MGDs exist, including more general definitions of the MGD properties, and well as different versions of the Lindeberg conditions (ii) and (ii'), see e.g. Brown (1971).

As an example of the application of the CLTs, consider:

Example I.4.13 *The empirical autocovariance function of order one for the ARCH(1) process* x_t*, is given by,*

$$\frac{1}{T}\sum_{t=1}^{T} x_t x_{t-1}. \tag{I.28}$$

If $\alpha < 1$*, such that* $\mathbb{E}[x_t^2] < \infty$*, the LLN indeed implies the obvious result that as* T *tends to* ∞*, then (I.28) will converge in probability to* $\mathbb{E}[x_t x_{t-1}] = 0$

using $g(x_t, x_{t-1}) = x_t x_{t-1}$. *Likewise, one would expect that multiplied by* $\sqrt{T}$, *the CLT in Theorem I.4.4 would apply to (I.28). Set therefore,*

$$Y_t = f(x_t, x_{t-1}) = x_t x_{t-1}.$$

Then Y_t *is a function of* (x_t, x_{t-1}) *and* $\mathbb{E}[Y_t | x_{t-1}] = 0$ *as desired. Moreover, if* $\mathbb{E}[x_t^4] < \infty$, *or* $\alpha < 1/\sqrt{3}$,

$$\mathbb{E}\left[Y_t^2\right] = \mathbb{E}\left[x_t^2 x_{t-1}^2\right] \leq \sqrt{\mathbb{E}[x_t^4]\,\mathbb{E}\left[x_{t-1}^4\right]} = \mathbb{E}\left[x_t^4\right] < \infty,$$

using Hölders inequality, $\mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}\left[|X|^2\right]\mathbb{E}\left[|Y|^2\right]}$ *for general random variables. Moreover, we can compute the variance,*

$$\mathbb{E}\left[x_t^2 x_{t-1}^2\right] = \mathbb{E}\left[x_{t-1}^2 \mathbb{E}\left[x_t^2 | x_{t-1}\right]\right] = \mathbb{E}\left[x_{t-1}^2\left(\sigma^2 + \alpha x_{t-1}^2\right)\right] = \mathbb{E}\left[x_{t-1}^2\right]\sigma^2 + \alpha\mathbb{E}\left[x_{t-1}^4\right],$$

with $\mathbb{E}\left[x_{t-1}^2\right]$ *and* $\mathbb{E}\left[x_{t-1}^4\right]$ *given in (I.19) and (I.20) respectively.*

We thus conclude that while $\alpha < 1$ *is sufficient for*

$$\frac{1}{T}\sum_{t=1}^{T} x_t x_{t-1} \xrightarrow{P} \mathbb{E}[x_t x_{t-1}] = 0,$$

we need the stronger assumption that $\alpha < 1/\sqrt{3}$ *for*

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

Often in applied work the empirical autocovariance function is studied for x_t^2, *given by,*

$$\frac{1}{T}\sum_{t=1}^{T}\left(x_t^2 - \left(\frac{1}{T}\sum_{t=1}^{T} x_t^2\right)\right) x_{t-1}^2. \tag{I.29}$$

Considerations as above lead to the requirement that $\mathbb{E}[x_t^4] < \infty$ *for convergence in probability, while by using Theorem I.4.3 the requirement is* $\mathbb{E}[x_t^8] < \infty$ *for convergence in distribution. These are quite strong restrictions, and therefore to avoid such, often the autocorrelation function is given for* $|x_t|$ *instead.*

I.5 Extending the AR(1)

A key process of interest is the vector version of the AR(1), the VAR(1), which for $X_t \in \mathbb{R}^p$, is given by

$$X_t = AX_{t-1} + \varepsilon_t, \quad t = 1, 2, ..., T \tag{I.30}$$

with initial value X_0 and ε_t i.i.d. $\mathrm{N}(0,\Omega)$, Ω symmetric and positive definite, $\Omega > 0$. Moreover, the autoregressive matrix $A = (A_{ij})_{i,j=1,2,\dots,p} \in \mathbb{R}^{p\times p}$, such that for $p = 2$, we have with $X_t = (X_{1t}, X_{2t})'$ and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$,

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{I.31}$$

or

$$\begin{aligned} X_{1t} &= A_{11}X_{1t-1} + A_{12}X_{2t-1} + \varepsilon_{1t} \\ X_{2t} &= A_{21}X_{1t-1} + A_{22}X_{2t-1} + \varepsilon_{2t} \end{aligned} \tag{I.32}$$

By definition of the multivariate Gaussian distribution, the transition density of $X_t \in \mathbb{R}^p$ conditional on X_{t-1} is given by

$$f(X_t|X_{t-1}) = \det(\Omega)^{-1/2} \left(1/\sqrt{2\pi}\right)^p \exp\left(-\frac{1}{2}(X_t - AX_{t-1})' \Omega^{-1} (X_t - AX_{t-1})\right)$$

which thus satisfies Assumption I.4.1. Moreover, as for the univariate case we have

$$X_t = \sum_{i=0}^{t-1} A^i \varepsilon_{t-i} + A^t X_0,$$

which, if $A^t \to 0$ as $t \to \infty$, resembles the stationary solution,

$$X_t^* = \sum_{i=0}^{\infty} A^i \varepsilon_{t-i}. \tag{I.33}$$

To see that a stationary solution exists, we may use the drift criterion with $\delta(X) = 1 + X'X = 1 + \|X\|^2$. Note first that,

$$X_{t+m} = \sum_{i=0}^{m-1} A^i \varepsilon_{t+m-i} + A^m X_t. \tag{I.34}$$

Recall that with $A \in \mathbb{R}^{p\times p}$, then $\mathrm{tr}(A) = \sum_{i=1}^{p} A_{ii}$, and hence by the properties of $\mathrm{tr}(\cdot)$, we have $\|X\|^2 = \mathrm{tr}(X'X) = \mathrm{tr}(XX')$, see e.g. Magnus and Neudecker (2007) for further properties and general results for matrix calculus. Using this and (I.34), we find

$$\begin{aligned} \mathbb{E}[\delta(X_{t+m})|X_t = X] &= 1 + \mathbb{E}\left[\mathrm{tr}\left(\sum_{i=0}^{m-1} A^i \varepsilon_{t+m-i}\varepsilon'_{t+m-i} \left(A^i\right)'\right)\right] + \|A^m X\|^2 \\ &= 1 + \underbrace{\mathrm{tr}\left(\sum_{i=0}^{m-1} A^i \Omega \left(A^i\right)'\right)}_{D_1} + \underbrace{\|A^m X\|^2}_{D_2} \end{aligned}$$

By Lemma A.1, in terms of the (sup-)matrix norm $\|\cdot\|$, we have

$$D_2 = \|A^m X\|^2 \leq \|A^m\|^2 \|X\|^2 .$$

With $\rho(A)$ denoting the spectral radius, that is, the largest (in absolute value) of the eigenvalues of A, we have $\|A^m\| \to 0$ as $m \to \infty$, provided $\rho(A) < 1$, see Lemma A.3. In particular this means that for $m > m^*$, say, if $\rho(A) < 1$,

$$\|A^m\| < \phi, \text{ with } \phi < 1.$$

Next, by Lemma A.4, as $\Omega > 0$ and $\rho(A) < 1$, with c a generic constant,

$$D_1 \leq \left\| \sum_{i=0}^{m-1} A^i \Omega \left(A^i\right)' \right\| \leq c \|\Omega\| \sum_{i=0}^{m-1} [\rho(A)]^i \to c \|\Omega\| (1 - \rho(A))^{-1} < \infty.$$

Hence we conclude that the drift criterion is satisfied with $\delta(X) = 1 + \|X\|^2$, as for $m > m^*$, and $\|X\|^2 > M$,

$$\mathbb{E}[\delta(X_{t+m}) | X_t = X] \leq g(X), \quad \text{with } g(X) = c + \phi \|X\|^2 ,$$

while $\mathbb{E}[\delta(X_{t+m}) | X_t = X] \leq c$ for $\|X\|^2 \leq M$, using Lemma A.4(iii) .

More generally, if $\rho(A) < 1$, the (unique) stationary solution is given by (I.33), and from properties of the Gaussian distribution, X_t^* have all moments finite.

Remark I.5.1 *Often in the time series literature, the condition* $\rho(A) < 1$ *is stated in terms of the so-called characteristic polynomial,* $A(z) = I_p - Az$, $z \in \mathbb{C}$. *Here the condition that* $\det(A(z)) = 0$ *implies* $|z| > 1$, *is equivalent to* $\rho(A) < 1$. *The condition is referred to as "the roots of the characteristic polynomial are larger than one in absolute value".*

I.5.1 Further extensions

Often one is interested in adding further lags, and consider therefore the VAR(k) process as given by

$$X_t = A_1 X_{t-1} + ... + A_k X_{t-k} + \varepsilon_t$$

with initial values $X_0, X_{-1}, ..., X_{-k}$ and ε_t i.i.d. $\mathrm{N}(0, \Omega)$. Obviously, in this case X_t does not satisfy Assumption I.4.1 as X_t by definition depends on $X_{t-1}, ..., X_{t-k}$. However, a way to circumvent this is to consider the so-called companion form in terms of

$$Y_t = \left(X_t', ..., X_{t-k-1}'\right)' \in \mathbb{R}^{pk}$$

which satisfies,

$$Y_t = AY_{t-1} + \epsilon_t,$$

where

$$A = \begin{bmatrix} A_1 & \cdots & A_{k-1} & A_k \\ I_p & & & \\ & \ddots & & \\ & & I_p & \end{bmatrix}$$

and $\epsilon_t = \left(\varepsilon_t', 0_p', ..., 0_p'\right)'$. Hence $Y_t \in \mathbb{R}^{pk}$ is indeed a Markov chain, has k-step transition density given by

$$\begin{aligned} f(Y_{t+k}|Y_t) &= f((X_{t+k}, ..., X_{t+1}) \,|\, (X_t, ..., X_{t-k})) \\ &= \frac{f(X_{t+k}, ..., X_{t-k})}{f(X_t, ..., X_{t-k})} \\ &= \prod_{m=1}^{k} f(X_{t+m}|X_{t+m-1}, ..., X_{t+m-k}), \end{aligned}$$

with $f(X_{t+m}|X_{t+m-1}, ..., X_{t+m-k})$ positive and continuous. Hence, as desired, Assumption I.4.1 holds for the Markov chain Y_t.

Application of the drift criterion gives, as for the VAR(1), that a sufficient condition for stationarity is indeed $\rho(A) < 1$. As for the VAR(1), this may alternatively be stated in terms of the characteristic polynomial which here is given by, $A(z) = I - \sum_{i=1}^{k} A_i z^i$, and the condition is as for the VAR(1), $\det(A(z)) = 0$, implies $|z| > 1$.

References

Brown, B.M. (1971) Martingale Central Limit Theorems, *The Annals of Mathematical Statistics.* 42:59-66.

Durrett, R. (2019) Probability Theory and Examples, 5th Edition, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

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

Horn, R.A. and Johnson, C.R. (2013) Matrix Analysis, Second Edition, Cambridge University Press.

Jensen, S.T., and Rahbek, A. (2007) On the Law of Large Numbers for (Geometrically) Ergodic Processes, *Econometric Theory*, 23:761-766.

Johansen, S. (1996) Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press.

Magnus, J. and Neudecker, H. (2007) Matrix Differential Calculus with Applications in Statistics and Econometrics, Third Edition, Wiley.

Meyn, S.P., and Tweedie, R.L. (1993) Markov chains and stochastic stability, Communications and Control Engineering Series, Springer-Verlag, London ltd., London.

Tjøstheim, D. (1990) Non-Linear Time Series and Markov Chains, *Advances in Applied Probability*, 22:587–611.

Tsay, R.S. (2010) Analysis of Financial Time Series, Wiley Series in Probability and Statistics, Third Edition, Wiley.

A Matrix results

Results on matrices and matrix norms can be found several places. The results listed here are from Horn and Johnson (2013).

Let M, N be real $p \times p$ matrices, then a matrix norm $\|\cdot\| : \mathbb{R}^{p \times p} \to \mathbb{R}$ satisfies

(i) $\|M\| \geq 0$

(ii) $\|M\| = 0$ if and only if $M = 0$

(iii) $\|aM\| = |a| \|M\|$ for $a \in \mathbb{R}$

(iv) $\|M + N\| \leq \|M\| + \|N\|$

(v) $\|MN\| \leq \|M\| \|N\|$.

Key examples of norms of matrices include the Euclidean norm given by,

$$\|M\|^2 = \operatorname{tr}(MM') = \sum_{i,j} |m_{ij}|^2,$$

and the "sup"-norm as given by,

$$\|M\| = \sup \{\|MX\| \mid X \in \mathbb{R}^p \text{ and } \|X\| \leq 1\}.$$

Here $\|X\|^2 = X'X$, $X \in \mathbb{R}^p$, and unless otherwise specified this (the Euclidean) norm for vectors will be used.

Lemma A.1 *With M a real $p \times p$ matrix, $\|\cdot\|$ the sup norm, and $X \in \mathbb{R}^p$,*

$$\|MX\| \leq \|M\| \|X\| \tag{I.35}$$

Lemma A.1 is stated in Theorem 5.6.2 in Horn and Johnson (2013) as a a simple consequence of the definition of the sup-norm.

For any $p \times p$ matrix M, whether symmetric or not, the eigenvalues λ_i, $i = 1, ..., p$, solve

$$\det(\lambda I_p - M) = 0$$

and the spectral radius $\rho(M)$ of M is given by

$$\rho(M) \equiv \max_{i=1,\dots,p} \{|\lambda_i|\} \tag{I.36}$$

This is an alternative measure of the 'size' of M and is related to any matrix norm by the following (Theorem 5.6.9 of Horn and Johnson, 2013) as follows:

Lemma A.2 *For any matrix norm* $\|\cdot\|$,

$$\rho(M) \leq \|M\|. \tag{I.37}$$

This reflects that M can have an arbitraily large norm, while the eigenvalues of M are small such as for

$$M = \begin{pmatrix} 0 & m_{12} \\ 0 & 0 \end{pmatrix}$$

with $\rho(M) = 0$, while $\|M\|$ can be made arbitraily large through m_{12}.

In the analysis of dynamic systems it is of interest to study iterations of the form $M^n X$ and the following results provide useful (Horn and Johnson, 2013, Theorem 5.6.12):

Lemma A.3 *With M a real $p \times p$ matrix, and any matrix norm* $\|\cdot\|$,

$$\lim_{n\to\infty} (M^n) = 0 \quad \textit{if and only if } \rho(M) < 1 \tag{I.38}$$

$$\rho(M) = \lim_{n\to\infty} \|M^n\|^{1/n} \tag{I.39}$$

Used frequently are the results addressing convergence of sums of M^n (Horn and Johnson, 2013, Lemma 5.6.10, see also Johansen, 1996, Corollary A.2):

Lemma A.4 *Assume that M a real $p\times p$ matrix, with eigenvalues $(\lambda_i)_{i=1,\ldots,p}$ and $\rho(M) < 1$. It then holds that:*

(i) *There exists a $c > 0$ such that for any $n \geq 0$, $\|M^n\| \leq c \cdot \rho(M)^{n/2}$*

(ii) $\lim_{n\to\infty} \left(\sum_{i=0}^{n} M^i\right) = \sum_{i=0}^{\infty} M^i = (I_p - M)^{-1}$

(iii) *With $\Omega > 0$, $\lim_{n\to\infty} \left(\sum_{i=0}^{n} M^i \Omega M^{i\prime}\right) = \sum_{i=0}^{\infty} M^i \Omega M^{i\prime} > 0$,*

with $\|\sum_{i=0}^{\infty} M^i \Omega M^{i\prime}\| < \infty$.

Proof: The results in *(ii)* and *(iii)* follow by Lemma A.3. To see this note that,

$$\sum_{i=0}^{n} M^i = \left(I_p - M^{n+1}\right)(I_p - M)^{-1}$$

since as $\rho(M) < 1$, all $|\lambda_i| < 1$. Hence *(ii)* holds by $M^{n+1} \to 0$. Likewise, for *(ii)*,

$$\left\|\sum_{i=0}^{n} M^i \Omega M^{i\prime}\right\| \leq \|\Omega\| \sum_{i=0}^{n} \left\|M^i\right\|^2 \leq c^2 \|\Omega\| \sum_{i=0}^{n} \rho(M)^{2i} \to c^2 \|\Omega\| \left(1 - \rho(M)^2\right)^{-1}.$$

The proof of *(i)* is more involved. Exploiting the Jordan form of M , cf. Lemma A.1 in Johansen (1996), gives that

$$\|M^n\| \leq \max_{i=1,\dots,p} |\lambda_i|^n P(n) \tag{I.40}$$

where $P(\cdot)$ is a polynomial of finite order (less than p). Hence, by

$$\|M^n\| \leq \rho(M)^n P(n) \leq \rho(M)^{n/2} \sup_n \rho(M)^{n/2} P(n) \leq c\rho(M)^{n/2} \tag{I.41}$$

the result in *(i)* follows. □