Anders Rahbek
Rasmus Søndergaard Petersen
University of Copenhagen

September 2022

Part I
ARCH(1) | An introduction to the theory of non-linear time series.

I.1 Introduction

In order to discuss econometric modeling of (conditional) time-varying volatility, or variance, this part introduces key ideas and theoretical probabilistic properties of the Autoregressive Conditional Heteroscedastic (ARCH) process of order one, ARCH(1).

The ARCH(1) is the simplest example of the rich class of ARCH models, and their generalized versions, the so-called GARCH models, which are dominating in empirical studies of (conditional) time-varying volatility in financial time series data. While the ARCH(1) process at a first glance appears simple, the probability analysis, as well as the statistical, or econometric analysis, of the ARCH(1) model, require general concepts from recent theory of non-linear models in statistics and econometrics.

The probability theory needed is introduced here such that it can also be applied in the context of the later studies and applications of the vast class of general non-linear ARCH volatility models and hence not alone for the ARCH(1).

To present the idea of modelling time-varying volatility, recall that in the option pricing literature a key assumption of the Black-Scholes model is that spot prices follow a geometric Brownian motion with constant drift and variance. That is, with $t = 1, 2, 3, ..$ denoting time and S_t the price of the underlying asset, log returns $r_t = \log(S_t) - \log(S_{t-1})$ are given by the equation,

$$r_t = \mu + \varepsilon_t,$$

where μ is the drift parameter and ε_t is identically and independently distributed (*i.i.d.*) following a Gaussian distribution with mean zero and variance

σ^2. Here σ (or, sometimes confusingly, σ^2) is commonly referred to as the 'volatility' which is, as is well-known, essential in option pricing and elsewhere. In this simple set-up of the classic Black-Scholes model there are two assumptions which are known from empirical analyses not to hold: *(i)* the (conditional) variance σ^2 is constant over time, and, *(ii)* log returns r_t are Gaussian distributed.

The class of ARCH processes, are processes for which the conditional volatility is allowed to be stochastic and time dependent - with periods of high conditional volatility and periods with low conditional volatility. Moreover, the tails of an ARCH-type process are more thick ('fat tails') than those of the Gaussian distribution. Thus the ARCH process seems a natural starting point for theoretical matching of empirically observed features of the data. Another typical feature of daily and weekly observed financial data series such as returns, is that the data in levels appear to be uncorrelated but not independent as the absolute value or the squared data series often appear empirically to be correlated. Again this is also reflected in the theoretical properties of the ARCH process.

Introductions to modelling issues of (G)ARCH models appear several places, see for example Taylor (2005) and Franses and van Dijk (2000, ch.3-4). A comparison of ARCH and other types of stochastic volatility models can be found in Shephard (1996). General overviews of the rich ARCH modelling literature are also found in e.g. Bollerslev, Chou and Kroner (1992), Bera and Higgins (1992) and Pagan (1996). Finally, the textbook by Francq and Zakoïan (2019) provides a rigorous mathematical statistical introduction to the theory of (G)ARCH models.

I.1.1 The ARCH(1) process

Consider initially the classic ARCH(1) model for (log returns) x_t, which for $t = 1, 2, \ldots$ can be represented as

$$x_t = \sigma_t z_t \tag{I.1}$$

$$\sigma_t^2 = \sigma^2 + \alpha x_{t-1}^2 \tag{I.2}$$

with initial value x_0 and where the z_t's are *i.i.d.* N(0,1). This way the distribution of x_t conditional on past information up to time $t-1$ as given by the variables $(x_{t-1}, ..., x_0)$ depends only on x_{t-1}. This is commonly referred to as the Markov property.

Moreover, the equations (I.1)-(I.2) imply that x_t conditionally on $(x_{t-1}, ..., x_0)$ is Gaussian distributed with conditional variance σ_t^2. However, *importantly, this does not imply that x_t are Gaussian distributed unconditionally.* Instead the marginal, or unconditional 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 a Gaussian distribution.

The level parameter σ^2 is strictly positive, $\sigma^2 > 0$ while $\alpha \geq 0$. Specifically if $\alpha = 0$ then x_t is simply an *i.i.d.* N(0,σ^2) sequence (conditionally and unconditionally).

It should be emphasized that in particular σ_t^2 is non-constant and stochastic.

The probabilistic behavior of the ARCH process x_t was until recently not fully described as a full discussion of for example the concept *stationarity* and *existence of moments* of x_t demands rather technical analysis. As indicated, the theory will be presented in such a way that other types of ARCH processes can be handled.

Specifically, 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 a simple restriction on the parameters for which x_t is well-behaved process in the sense that it is *stationary* and *asymptotically independent (weakly mixing)* and hence satisfying the *law of large numbers* – which again is important for a discussion of estimation and inference in ARCH and other econometric volatility models.

The concepts *stationarity, asymptotic independence* and *the law of large numbers* will be given precise meaning and definitions in the following sections.

Note that the specification of σ_t^2 is referred to as the linear ARCH(1) model as it is linear in x_{t-1}^2 and depends only on one lag of x_t. The linearity assumption is empirically questionable as for example responses to changes in x_{t-1}^2 often appear nonlinear and asymmetric. In the above mentioned surveys of ARCH modelling much more general functional forms of σ_t^2 are introduced . We shall return to some of these and the ideas behind them later but introduce the theory in terms of the linear ARCH process, simply referred to as the ARCH process henceforth unless otherwise stated. In addition to the surveys, classic references for ARCH models are Engle (1982) and Nelson (1990).

I.2 Conditional moments

Here some key conditional and unconditional moments of the ARCH process are derived. Conditional expectations play a key role in these considerations and will therefore be discussed briefly. Throughout this section we will work under the assumption that all moments, conditional as well as unconditional, are well-defined such that the calculations are valid. Note that if x_t is *i.i.d.* Gaussian, then all moments of x_t are by definition finite, that is,

$$E|x_t|^k < \infty,$$

for any $k \geq 0$. As will be demonstrated this is *not* the case for non-Gaussian distributions such as for example the ARCH process. In the next section we do establish under which assumptions on the two parameters σ^2 and α the different moments are well-defined and finite.

First we need to give meaning to for example the conditional expectation $E(x_t|x_{t-1})$ and be able to work with conditional expectations in general.

I.2.1 Conditional expectations

Consider in general two random, or equivalently stochastic, variables X and Y, with X having finite expectation $E|X| < \infty$. We shall work with real valued stochastic variables, $X \in \mathbb{R}$, and vector valued, $X \in \mathbb{R}^p$ for some $p \geq 1$ and likewise for Y. We shall furthermore work under the assumption that X has density $f(x)$, Y has density $f(y)$ and the conditional distribution of X given Y has density $f(x|y)$.

In terms of the density, recall that the expectation of X, $X \in \mathbb{R}$ or $X \in \mathbb{R}^p$, can be computed as

$$E(X) = \int_{\mathbb{R}^p} x f(x)\, dx. \tag{I.3}$$

Likewise, we define the conditional expectation of X given $Y = y$, denoted $E(X|Y = y)$, by

$$E(X|Y = y) = \int_{\mathbb{R}^p} x f(x|y)\, dx, \tag{I.4}$$

which, by definition, depends on the value y.

Example I.2.1 *If X is $N(\mu_x, \sigma_x^2)$ distributed, then the density of X is given by,*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{1}{2\sigma_x^2}(x-\mu_x)^2\right),$$

and from probability analysis, $E(X) = \mu_x = \int_{\mathbb{R}} x f(x)\, dx$.

Example I.2.2 *Consider $(X, Y)'$ which is bivariate $N_2(\mu, \Omega)$ distributed, with mean $\mu = E(X, Y)'$ and variance matrix $\Omega = Var(X, Y)'$ given by,*

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \quad \Omega = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix}.$$

In particular, X *is* $N(\mu_x, \sigma_x^2)$ *and* Y *is* $N(\mu_y, \sigma_y^2)$ *distributed, and* $Cov(X, Y) = \sigma_{xy}$.

Recall the result that the conditional expectation of X *given* $Y = y$, $E(X|Y = y)$*, is given by*

$$E(X|Y = y) = \mu_x + \omega(y - \mu_y) = \mu_{x|y}, \quad \textit{where } \omega = \sigma_{xy}/\sigma_y^2. \tag{I.5}$$

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

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

and hence the conditional density is given by

$$\begin{aligned} f(x|y) &= \frac{1}{\sqrt{2\pi\sigma_{x|y}^2}} \exp\left(-\frac{1}{2\sigma_{x|y}^2}\left(x - \mu_{x|y}\right)^2\right) \\ &= \frac{1}{\sqrt{2\pi\sigma_{x|y}^2}} \exp\left(-\frac{1}{2\sigma_{x|y}^2}\left(x - \mu_x - \omega(y - \mu_y)\right)^2\right). \end{aligned}$$

Thus as illustrated in Example I.2.2, by defining the conditional expectation $E(X|Y = y)$ as in (I.4), for each value of y we get a different value. In general, we wish to use the conditional expectation for all possible values of Y, that is to treat the conditional expectation as a random variable which we call $E(X|Y)$. With $g(y) = E(X|Y = y)$ defined in (I.4), this is accomplished by simply setting,

$$E(X|Y) = g(Y). \tag{I.6}$$

This is a function of Y and therefore a random variable and defines $E(X|Y)$.

Example I.2.3 *In terms of Example I.2.2, then*

$$E(X|Y) = \mu_x + \omega(Y - \mu_y).$$

This is a random variable (it is a function of Y*), and moreover it has the same expectation as* X *since*

$$E(E(X|Y)) = E(\mu_x + \omega(Y - \mu_y)) = \mu_x + \omega(E(Y) - \mu_y) = \mu_x = E(X). \tag{I.7}$$

The fact that $E(E(X|Y)) = E(X)$ in the above example is a general feature of the conditional expectation, often referred to as the law of iterated expectations. This is a much used implication of the defintion of conditional expectations. Before listing some rules much useful for calulations with conditional expectations we note that if the joint distribution of X and Y has density $f(x,y)$, the conditional density of X given Y is given by

$$f(x|y) = f(x,y)/f(y). \tag{I.8}$$

This corresponds to the well-known result that

$$P(X \in A|Y \in B) = P(X \in A, Y \in B)/P(Y \in B).$$

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 $E(X|Y)$ it holds that $E|E(X|Y)| < \infty$ and the law of iterated expectations apply,*

$$E(X) = E(E(X|Y)). \tag{I.9}$$

If X and Y are independent,

$$E(X|Y) = E(X). \tag{I.10}$$

Moreover,

$$E(X|Y) = E(E(X|Y,Z)|Y) \qquad \textit{and} \tag{I.11}$$

$$E(X|X) = X \tag{I.12}$$

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

$$E(g(Y) + h(Y)X|Y) = g(Y) + h(Y)E(X|Y). \tag{I.13}$$

Example I.2.4 *Consider the AR process given by,*

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

for $t = 1, 2, ..., T$, initial value x_0 and where ε_t is i.i.d.N(0,σ^2). Then

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

where we have used (I.13), (I.12) and (I.10).

Example I.2.5 *Consider the ARCH process x_t in (I.1), then using first (I.13) and next (I.10) we find that the conditional expectation of x_t is zero as,*

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

Hence the ARCH(1) process has mean zero as $E\left(x_t\right)=E\left(E\left(x_t|x_{t-1}\right)\right)=0$ by (I.9).

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

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

such that,

$$
E\left(X|X\in A\right)=\int xf\left(x|A\right)dx.
$$

Example I.2.6 *With x_t $N\left(0,\sigma^2\right)$, distributed, the density of the distribution of x_t conditional on $x_t>0$ is (using $P\left(x_t>0\right)=\frac{1}{2}$) given by*

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

Next, with z_t $N\left(0,1\right)$ distributed,

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

using well-known properties of the Gaussian distribution: (i) symmetry, and (ii) $E|z_t|=\sqrt{2/\pi}$.

I.2.2 The ARCH(1) process continued

We discuss here some first properties of the ARCH(1) process x_t defined in (I.1). It should be emphasized that all calculations in the next are done under the assumption of relevant finite order moments of the ARCH process which we will establish in the next section.

It is useful to discuss conditioning not only only x_{t-1} as in Example I.2.5 but also on all past variables $(x_{t-1}, x_{t-2}..., x_0)$. This is often referred to as 'conditioning on all past information' in the literature.

In terms of the definition of conditional expectations, we give meaning to $E(x_t|x_{t-1}, ..., x_0)$ by setting $X = x_t$ and $Y = (x_t, x_{t-1}, ..., x_0)$, and write it as

$$E(x_t|\mathcal{F}_{t-1}),$$

with

$$\mathcal{F}_{t-1} = (x_{t-1}, x_{t-2} \ldots, x_0).$$

In terms of this notation we can as in the example find the first moment of the ARCH process x_t,

$$E(x_t) = E(E(x_t|\mathcal{F}_{t-1})) = E(\sigma_t E(z_t)) = 0.$$

That the calculations are identical, reflects that by the defintion of the ARCH(1) process, the distribution of x_t conditional on $(x_{t-1}, x_{t-2}, ..., x_0)$ depends only on x_{t-1}.

Consider next the correlation between x_t and x_{t-1}, say

$$E(x_{t-1}x_t) = E(E(x_t x_{t-1}|\mathcal{F}_{t-1})) = E(x_{t-1}E(x_t|\mathcal{F}_{t-1})) = 0$$

and likewise,

$$E(x_{t-k}x_t) = E\left(E\left(x_{t-k}x_t|\mathcal{F}_{t-k}\right)\right) = 0 \text{ for any } k \geq 1$$

Hence the ARCH(1) process is a mean zero and uncorrelated process.

Note that this – unlike for Gaussian variables – *does not imply* that x_t and x_{t-1} are independent as σ_t^2, and hence x_t, depends on x_{t-1}^2.

Now turn to the second order moment, or as $Ex_t = 0$, the variance:

$$V(x_t) = E(x_t^2) = E\left(E\left(x_t^2|\mathcal{F}_{t-1}\right)\right) = E(\sigma_t^2 E(z_t^2)) = E(\sigma_t^2) = \sigma^2 + \alpha E\left(x_{t-1}^2\right) \tag{I.14}$$

Hence, if $E(x_t^2)$ is constant and finite (or, $\alpha < 1$), the variance of x_t is given by,

$$V(x_t) = E\left(x_t^2\right) = \frac{\sigma^2}{1-\alpha}. \tag{I.15}$$

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

$$E(x_t^{2k+1}) = 0$$

i.e. all odd moments are zero. For the 4th order moment it follows that,

$$E(x_t^4) = 3(\tfrac{1-\alpha^2}{1-3\alpha^2})E(x_t^2)^2 > 3E(x_t^2)^2 \tag{I.16}$$

if $\alpha < 1/\sqrt{3}$. As $E(x_t^4)/E(x_t^2)^2 = 3$ for the Gaussian case, we conclude that the ARCH(1) process has indeed fatter tails or so-called *excess kurtosis* since $1-\alpha^2 > 1-3\alpha^2$, or $\alpha < 1/\sqrt{3}$.

I.3 Stationarity and mixing

Above we computed moments $E\left(x_t^k\right)$ for the ARCH(1) process assuming that they were the same for all time points, that is $E\left(x_t^k\right) = E\left(x_{t+n}^k\right)$ for all n, and that they were finite. To show under which conditions on the parameters this is the case we introduce theory from probability analysis. We first discuss stationarity.

I.3.1 Stationarity

A time series or, equivalently, a stochastic process, $(X_t)_{t=0,1,2,\ldots}$ is a sequence of stochastic variables with each X_t taking values in $\mathbb{R}$ or $\mathbb{R}^p$. A key concept in time series analysis is that of stationarity of a stochastic process which can be formally defined as follows:

Definition I.3.1 *The process $(X_t)_{t=0,1,2,\ldots}$ is said to be stationary, or simply X_t is said to be stationary, if for all $h, h \geq 0$, and t 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 $E(X_t)$ and variance $V(X_t)$ are constant, while the covariance between X_t and X_{t+h}, $Cov\left(X_t, X_{t+h}\right)$ depends only on h, and not on t.

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

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

with $\Omega_h = \sigma^2 I_{h+1}$ *where* I_{h+1} *is the* $(h+1)$*-dimensional identity matrix. We can also write*

$$\Omega_h = diag\left(\sigma^2, ..., \sigma^2\right).$$

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, $E(X_t)$ is constant and $Cov(X_t, X_{t+h}) = v(h)$ 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.* $N(0, \sigma^2)$*. In particular* x_t *is a stationary process with* $Ex_t = 0$*,* $V(x_t) = (1+\theta^2)\sigma^2$*,* $Cov(x_t, x_{t+1}) = \theta\sigma^2$ *and*

$$Cov(x_t, x_{t+h}) = 0 \quad \text{for } h > 1.$$

Hence the MA(1) process is stationary.

In the next example we use the result:

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*

$$\sum_{i=0}^{\infty} \phi^i = 1/(1-\phi). \tag{I.17}$$

Example I.3.3 *Consider the AR process given by,*

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

for $t = 1, 2, ..., T$*, initial value* x_0 *and where* ε_t *i.i.d.N(0,*σ^2*). Simple recursion shows that*

$$x_t = \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i} + \rho^t x_0. \tag{I.18}$$

Hence in particular,

$$E(x_t) = \rho^t x_0$$

which depends on t *and the AR process is thus not stationary. However, we can make it stationary by choosing* x_0 *as below and restricting* ρ *such that* $|\rho| < 1$*. In this case* x_t *has the stationary distribution,*

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

To see this, we can give the initial value x_0 *the distribution of* x_0^**, that is* $x_0 = \sum_{i=0}^{\infty} \rho^i \varepsilon_{0-i}$*. Then simple insertion in (I.18) indeed gives,*

$$x_t = \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i} + \rho^t \sum_{i=0}^{\infty} \rho^i \varepsilon_{0-i} = \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i} + \sum_{i=t}^{\infty} \rho^i \varepsilon_{t-i} = x_t^*.$$

And as (x_t^*) *is Gaussian with* $E(x_t^*) = 0$*, using (I.17) with* $\phi = \rho^2$*,*

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

and using that ε_t *are i.i.d. such that* $Cov(\varepsilon_t \varepsilon_{t+h}) = 0$ *for* $h \geq 1$*, we also find the formula for the covariance of a stationary AR(1) process,*

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

we conclude that x_t^* *is stationary.*

That is, if $|\rho| < 1$ *the initial value* x_0 *can be given an initial distribution such that* $x_t = x_t^*$ *is 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 theory below will show that it is in fact not needed to introduce the inifinite sums and the explicit stationary solution* x_t^* *to discuss stationarity and dependence structure of the AR process.*

More generally, we can apply the concept of stationarity to the familiar autoregressive (AR) and moving average (MA) processes by considerations as above. However, we cannot do so for the ARCH processes and different techniques will be applied. These techniques can of course also be applied to AR and MA processes as noted in the previous example.

I.3.2 Weakly mixing and asymptotic independence

The definition of stationarity adresses 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. To measure correlation over time for stationary processes usually the covariance function,

$$v(h) \equiv Cov(X_t, X_{t+h}), \tag{I.19}$$

is considered. As already noted for a stationary process this does not depend on t. For a univariate stationary process the well-known autocorrelation function is defined by,

$$\rho(h) = Corr\left(X_t, X_{t+h}\right) = \frac{Cov(X_t, X_{t+h})}{\sqrt{V\left(X_t\right) V\left(X_{t+h}\right)}} = \frac{v(h)}{v(0)}, \tag{I.20}$$

where the last equality holds by stationarity.

The functions $\rho(h)$ and $v(h)$ for various h describe the correlatedness over periods of time. Weakly mixing, or asymptotic indepencence, which is defined next, states that the dependence between X_t and X_{t+h} vanishes as h increases and hence replaces independence. More precisely:

Definition I.3.2 *A stationary process* $(X_t)_{t=0,1,\ldots}$ *is said to be weakly mixing, if for all* t, h *and sets* A, B,

$$\begin{aligned} &P((X_0, \ldots, X_t) \in A,\ (X_h, \ldots, X_{t+h}) \in B) \\ &\to P((X_0, \ldots, X_t) \in A) P((X_0, \ldots, X_t) \in B) \end{aligned}$$

as $h \to \infty$.

Intuitively this means that events removed from one another in time are independent. The next result relates correlation with weakly mixing:

Lemma I.3.2 *If the stationary process* $(X_t)_{t=0,1,2,..}$ *is weakly mixing and has finite variance, then the covariance* $v(h) = Cov(X_t, X_{t+h})$ *tends to zero as* $h \to \infty$.

If X_t *is a stationary Gaussian process and* $v(h) \to 0,\ h \to \infty$ *then* X_t *is weakly mixing.*

Example I.3.4 *It follows by Example I.3.2 that the MA(1) process is weakly mixing and likewise, if* $|\rho| < 1$, *the AR(1) process is weakly mixing by Example I.3.3.*

A key implication of mixing is that a law of large numbers (LLN) apply:

Theorem I.3.1 *Assume that with* $X_t \in \mathbb{R}^p$, $(X_t)_{t=0,1,2,3,...}$ *is a weakly mixing process, and assume that the function* $g : \mathbb{R}^{p(m+1)} \to \mathbb{R}$, $m \geq 0$, *satisfies* $E|g(X_t, X_{t+1}, ..., X_{t+m})| < \infty$.*Then as* $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T} g(X_t, X_{t+1}, ..., X_{t+m}) \xrightarrow{P} E(g(X_t, X_{t+1}, ..., X_{t+m})). \tag{I.21}$$

This version of the LLN applies a quite general formulation in terms of the function g.

Example I.3.5 *The formulation means that for example if* X_t *is univariate and mixing with finite second order moments, then*

$$\frac{1}{T}\sum_{t=1}^{T} X_t^2 \xrightarrow{P} E\left(X_t^2\right) \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T} X_t X_{t+1} \xrightarrow{P} E(X_t X_{t+1}),$$

by applying Theorem I.3.1 with

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

Example I.3.6 *If* $EX_t = 0$, *such that* $Var(X_t) = E(X_t^2)$ *and* $Cov(X_t, X_{t+h}) = E(X_t X_{t+h})$ *it follows that the empirical autocorrelation function, see (I.20), as defined by,*

$$\widehat{\rho}(h) \equiv \frac{\frac{1}{T}\sum_{t=1}^{T-h} X_t X_{t+h}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} X_t^2 \frac{1}{T}\sum_{t=1}^{T-h} X_{t+h}^2}}, \tag{I.22}$$

will converge in probability to (the theoretical) $\rho(h)$ *by Theorem I.3.1 motivating that most software for time series allows one to compute these directly.*

I.3.3 Moments and stationarity using the drift criterion

Before turning to the concept of a drift function and the drift criterion, we note the key implication that if $(X_t)_{t=0,1,2,...}$ satifies the drift criterion, the initial value, X_0, can be given a distribution such that indeed X_t is stationary. This resembles the considerations we made for the AR(1) process x_t in Example I.3.3 where we could explicitly choose an initial distribution of x_0 such that x_t became stationary. Moreover, with X_0 given this initial distribution, the stationary process is also *weakly mixing* such that the law of large numbers can be applied. And in addition, the drift criterion, implies that X_t has finite moments as we will see.

Hence establishing the drift criterion is very helpful in many ways. The presentation here is based on Meyn and Tweedie (1993), Tjøstheim (1990) and Hansen and Rahbek (1998), see alsoCarrasco and Chen (2002) for general ARCH and related stochastic volatility processes.

I.3.3.1 Assumptions

Now a common key feature of the the AR(1) process in Example I.2.4 and the ARCH(1) process in (I.1) is that with X_t denoting either of the two, then the conditional distribution of,

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

depends only on X_{t-1}. More precisely, in the AR(1) case x_t conditionally on x_{t-1}, is $N(\rho x_{t-1}, \sigma^2)$ distributed, while for the ARCH(1) x_t conditionally on x_{t-1}, is $N(0, \sigma_t^2)$ 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 attractable features. We make the following assumption:

Assumption I.3.1 *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-1}, has a positive conditional density $f(X_t|X_{t-1})$, $f(X_t|X_{t-1}) > 0$, which is continuous.*

Note that (i) implies that $(X_t)_{t=0,1,2,\ldots}$ is a Markov chain on $\mathbb{R}^p$, that is a Markov chain with non-integer values, or what is called a Markov chain on a general state space. The condition (ii) of continuity, while simple to validate, is not needed and milder conditions on the conditional density can be applied in general.

Example I.3.7 *For the AR process x_t in Example I.2.4, x_t conditional on x_{t-1} has density*

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

which is positive and continuous in x_t and x_{t-1}.

Example I.3.8 *For the ARCH process in (I.1), 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 again is continuous as desired.

We shall in the next use the follwing result from probability analysis:

Lemma I.3.3 *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 $VX < \infty$, then $EY = cEX$ and $VY = c^2 VX$.*

Example I.3.9 *In the ARCH process x_t in (I.1) the assumption of z_t i.i.d.$N(0,1)$ is sometimes replaced by the assumption that z_t is i.i.d.$D(0,1)$ where D is a 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.3.1.*

To see this note first that if X is t_v-distributed with $v > 2$, then X has $EX = 0$ and $VX = v/(v-2)$. 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 $\gamma(v) = \Gamma\left(\frac{v+1}{2}\right) / \sqrt{\Gamma\left(\frac{v}{2}\right)}$, with $\Gamma(\cdot)$ the so-called Gamma function.

As $VX = v/(v-2)$, then using Lemma I.3.3, $z_t = \left(\sqrt{\frac{v-2}{v}}\right) X$ has $V(z_t) = 1$ and $E(z_t) = 0$, explaining the assumption on the i.i.d.D(0,1) innovations z_t. It follows likewise by simple insertion that 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 hence, using Lemma I.3.3 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.3.3.2 Drift function

Next, we define the concept of *a drift function* for a process X_t satisfying Assumption I.3.1. With X_t a time series, a drift function for X_t is some function δ, 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.3.10 *A much used drift function in the analysis of univariate AR and ARCH processes is*

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

If X_t is a vector, $X_t = (X_{1t}, ..., X_{pt})' \in \mathbb{R}^p$, then one may apply,

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

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. We do this by considering the conditional expectation of $\delta(X_t)$ given X_{t-1} or some other past value of X_t, say X_{t-m}. That is we are interested in measuring

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

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

Example I.3.11 *Consider the AR(1) process x_t in Example I.2.4 and set $\delta(x_t) = 1 + x_t^2$. Then*

$$\begin{aligned}
E(\delta(x_t)|x_{t-1}) &= E\left(1 + (\rho x_{t-1} + \varepsilon_t)^2 |x_{t-1}\right) \\
&= 1 + \rho^2 E\left(x_{t-1}^2|x_{t-1}\right) + 2\rho x_{t-1} E(\varepsilon_t|x_{t-1}) + E\left(\varepsilon_t^2|x_{t-1}\right) \\
&= 1 + \rho^2 x_{t-1}^2 + 2\rho x_{t-1} E(\varepsilon_t) + E\left(\varepsilon_t^2\right) \\
&= 1 + \sigma^2 + \rho^2 x_{t-1}^2 \\
&= \rho^2 \delta(x_{t-1}) + c, \qquad \text{(I.23)}
\end{aligned}$$

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 $\eta_t = (\delta(x_t) - E(\delta(x_t)|x_{t-1}))$ and such that by defintion $E\eta_t = 0$. The equation for $\delta(x_t)$ is stable if $\rho^2 < 1$ and persistent if $\rho = 1$. That is, the dynamics of the drift function $\delta(x_t)$ reflects the dynamics of the AR process x_t.

Example I.3.12 *Consider the ARCH process x_t given by (I.1) and set $\delta(x_t) = 1 + x_t^2$. Then*

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

where $c = (1 + \sigma^2 - \alpha)$. Thus as before in the AR example, we can interprete this as a simple autoregression in $\delta(x_{t-1})$ with autoregressive coefficient α. This is stable if $\alpha < 1$ (as $\alpha \geq 0$ by definition of the ARCH process). That indeed the ARCH process x_t itself is a stable and (asymptotically) stationary process if $\alpha < 1$ is a key implication of the next.

In the above examples the dynamics of the drift function, or rather the conditional expectation of $\delta(X_t)$ given X_{t-m}, we used the concept of stability informally. More precisely, we shall make the following assumption:

Assumption I.3.2 *Assume that $(X_t)_{t=0,1,2,..}$, with $X_t \in \mathbb{R}^p$, satifies Assumption I.3.1. Assume further that there is a drift function δ, $\delta(X_t) \geq 1$,* *which for some lag m, there are positive constants M, C and ϕ, $\phi < 1$, such that,*

$$\begin{array}{lll} (i) & E\left(\delta\left(X_t\right)|X_{t-m}\right) \leq \phi\delta\left(X_{t-m}\right) & \text{for } X_{t-m}'X_{t-m} > M, \\ (ii) & E\left(\delta\left(X_t\right)|X_{t-m}\right) \leq C & \text{for } X_{t-m}'X_{t-m} \leq M. \end{array}$$

If (X_t) satisfies Assumption I.3.2 then we say that X_t satisfies the drift criterion (with drift function δ).

Example I.3.13 *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. To see this note first that the calculations in Example I.3.10 gave,*

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

As $\rho^2 < 1$ we may choose ϕ smaller than 1, but larger than ρ^2. That is, we choose ϕ such that $\rho^2 < \phi < 1$. With this choice of ϕ, $(\phi - \rho^2) > 0$, and hence for x_{t-1}^2 large enough, $x_{t-1}^2 > M$ say,

$$c < \left(\phi - \rho^2\right)\delta\left(x_{t-1}\right).$$

We can therefore conclude that there for some large $M > 0$, and $x_{t-1}^2 > M$,

$$E\left(\delta\left(x_t\right)|x_{t-1}\right) = \rho^2\delta\left(x_{t-1}\right) + c < \phi\delta\left(x_{t-1}\right).$$

For $x_{t-1}^2 \leq M$, we have $\rho^2\delta\left(x_{t-1}\right) + c \leq \rho^2\delta\left(M\right) + c = C$.

Example I.3.14 *The ARCH process x_t satisfies the drift criterion if $\alpha < 1$ with $\delta(x_{t-1}) = 1 + x_{t-1}^2$. Using that by Example I.3.12,*

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

the considerations in Example I.3.13 carry over directly with $\alpha = \rho^2$.

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

Theorem I.3.2 *Assume that $(X_t)_{t=0,1,\ldots}$ satisfies the drift criterion with drift function δ. Then the initial value, X_0, can be given a distribution such that $(X_t)_{t=0,1,2}$ is a stationary process. Moreover, the stationary process is weakly mixing and has finite moments, $E\left(\delta\left(X_t\right)\right) < \infty$.*

Finally, for any initial value X_0, the law of large numbers in (I.21) in Theorem I.3.1 applies to X_t.

Thus, if X_t satisfies the drift criterion, not only is the process stationary (by giving X_0 the correct distribution) and weakly mixing such that the law of large numbers apply, but as $E(\delta(X_t)) < \infty$, then any moments of X_t which are bounded by the drift function δ are finite.

Furthermore, the last statement in the theorem says that if X_0 does not have the correct stationary initial distribution, the importance of this will vanish. In fact, it will vanish exponentially fast, and the distribution of X_t for moderately large t will resemble the stationary distribution. This feature is often referred to as *geometric ergodicity.*

Example I.3.15 *By Example I.3.13 For the AR process we may conclude that $Ex_t^2 < \infty$, and that the law of large numbers apply to the stationary x_t by Theorem I.3.2. 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 and since it is Gaussian actually $Ex_t^{2k} < \infty$ for any k. To give an understanding of the role of initial value, recall from Example I.3.3 that with x_0 fixed,*

$$x_t = \rho^t x_0 + \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i} \stackrel{D}{=} N\left(\rho^t x_0, \sigma^2 \frac{1-\rho^{2t}}{1-\rho^2}\right),$$

while for the stationary version,

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

We observe that $\rho^t x_0 \to 0$ exponentially fast, and likewise $\frac{1-\rho^{2t}}{1-\rho^2} \to \frac{1}{1-\rho^2}$, and hence the initial value plays no role asymptotically.

Example I.3.16 *For the ARCH 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 may from Example I.3.14 conclude that if $0 \le \alpha < 1$ then x_t has a stationary solution with $Ex_t^2 < \infty$. Hence any moments of order lower than 2 are finite, for example $E|x_t| < \infty$ since $|x_t| \le \delta(x_t) = 1 + x_t^2$.

We do not know if for example x_t has finite fourth order moments, $Ex_t^4 < \infty$. To find out under which conditions this holds we need to consider a drift function from which we can conclude this. An example is

$$\delta(x_t) = 1 + x_t^4.$$

With z_t i.i.d.$N(0,1)$, $E(z_t^4) = 3$ and we find,

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

where $c\left(x_{t-1}^2\right) = c + 6\alpha\sigma^2 x_{t-1}^2$, with $c = (1 - 3\alpha^2 + 3\sigma^4)$. We thus need to choose α so small that $3\alpha^2 < 1$. To see that the drift-criterion holds with such an α, choose ϕ such that $3\alpha^2 < \phi < 1$. Next we need to establish that for for some large M and $x_{t-1}^2 > M$,

$$\left(\phi - 3\alpha^2\right)\delta(x_{t-1}) > c\left(x_{t-1}^2\right).$$

But this clearly holds as $x_{t-1}^4 > x_{t-1}^2$ for large x_{t-1}^2 and $(\phi - 3\alpha^2) > 0$.

Hence the conclusion is that while a stationary x_t exists for $\alpha < 1$ and $Ex_t^2 < \infty$ 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 $Ex_t^4 < \infty$.

We saw in Example I.3.16 that the value of α in the conditional variance $\sigma_t^2 = \sigma^2 + \alpha x_{t-1}^2$ was cruical for stationarity of x_t and also for 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.

In terms of the AR process x_t the autoregressive coefficient is the key parameter to an understanding of the dynamics of $x_t = \rho x_{t-1} + \varepsilon_t$ as is well-known. We know that if $|\rho| < 1$ then x_t is stationary and with ε_t Gaussian all moments are finite. At the same time we know that the restriction of $|\rho| < 1$ is cruical in the sense that if $\rho = 1$, as often found empirically, then x_t is a unit-root process and non-stationary. Surprisingly this is not so for the ARCH process, where $\alpha = 1$ still allows x_t to be stationary. However, with $\alpha = 1$, then only moments up to order one are finite, $E|x_t| < \infty$. That is, $\alpha = 1$ implies stationarity but for a process without 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.1):	
$x_t = \sigma_t z_t,\ \ \sigma_t = \sigma^2 + \alpha x_{t-1}^2$ and z_t $i.i.d.\mathrm{N}(0,1)$.	
Stationary for $0 \leq \alpha < 3.56$	
Finite moments:	
$\frac{\pi}{2} \leq \alpha < 3.56$	no moments (fractional)
$1 \leq \alpha < \frac{\pi}{2}$	$E\|x_t\| < \infty$
$\frac{1}{\sqrt{3}} \leq \alpha < 1$	$Ex_t^2 < \infty$
$0 \leq \alpha_1 < \frac{1}{\sqrt{3}}$	$Ex_t^4 < \infty$

Actually, see Nelson (1990), the 3.56 is an approximation of the number $\frac{1}{2}\exp(-\Psi(\frac{1}{2}))$ where $\Psi(\cdot)$ is the Euler psi function with $\Psi(\frac{1}{2}) \cong -1.96351$. The number appears from considering the condition for stationarı́ty

$$E\left(\log\left(\alpha z_t^2\right)\right) < 0,$$

which can be solved for z_t $i.i.d.\mathrm{N}(0,1)$ as above. In fact, if z_t has another distribution such as the t_v distribution mentioned in Example I.3.9 the intervals above would change.

This as well as other examples of ARCH models will be discussed in exercises.

I.4 Central limit theory

As noted, a powerful implication of Theorem I.3.2 for X_t is that independently of the initial value, X_0 the LLN in Theorem I.3.1 applies. That is,

$$\frac{1}{T}\sum_{t=1}^{T} g\left(X_t, X_{t+1}, ..., X_{t+m}\right) \xrightarrow{P} E\left(g\left(X_t, X_{t+1}, ..., X_{t+m}\right)\right), \tag{I.24}$$

provided that $E|g\left(X_t, ..., X_{t+m}\right)| < \infty$.

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:

Theorem I.4.1 (Meyn and Tweedie, 1993, Theorem 17.0.1) *Assume that Theorem I.3.2 applies to* $(X_t)_{t\geq 0}$ *with* X_t *stationary. With* $f\left(X_t, X_{t-1}, ..., X_{t-m}\right) \in \mathbb{R}$*, assume that* $E|f^2\left(X_t, ..., X_{t-m}\right)| < \infty$*, and* $Ef(X_t, X_{t-1}, ..., X_{t-m})| = 0$*. Then,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(X_t, X_{t-1}, ..., X_{t-m}) \xrightarrow{D} N(0,\gamma),$$

where

$$\gamma = \lim_{T\to\infty} E\left[\frac{1}{T}\left(\sum_{t=1}^{T} f(X_t, X_{t-1}, ..., X_{t-m})\right)^2\right].$$

A different version of the CLT is from Brown (1971) for *Martingale differences* which is the one we will apply when discussing asymptotic normality later for estimators[1]:

Theorem I.4.2 (Corollary to Brown, 1971) *Assume that Theorem I.3.2 applies to $(X_t)_{t\geq 0}$ with X_t stationary. With $f\left(X_t, X_{t-1}, ..., X_{t-m}\right) \in \mathbb{R}$, assume that $E|f^2\left(X_t, ..., X_{t-m}\right)| < \infty$, and $E\left(f(X_t, X_{t-1}, ..., X_{t-m})|X_{t-1}, ..., X_{t-m}\right) = 0$. Then the central limit theorem (CLT) applies as $T \to \infty$,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(X_t, X_{t-1}, ..., X_{t-m}) \xrightarrow{D} N(0, E\left(f^2(X_t, X_{t-1}, ..., X_{t-m})\right)).$$

As an example of the application of the CLT quoted, consider:

Example I.4.1 *The empirical autocovariance function of order one for the ARCH(1) process x_t, is given by,*

$$\frac{1}{T}\sum_{1}^{T} x_t x_{t-1}. \tag{I.25}$$

If $\alpha < 1$, such that $Ex_t^2 < \infty$, the LLN indeed implies the obvious result that as T tends to ∞, then (I.25) will converge in probability to $Ex_t x_{t-1} = 0$ using $g\left(x_t, x_{t-1}\right) = x_t x_{t-1}$.

Likewise, one would expect that multiplied by $\sqrt{T}$, the CLT in Theorem I.4.2 would apply to (I.25). Set therefore,

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

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

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

[1]Note that under an additional regularity condition (often referred to as "Lindeberg") then the stationarity requirement may be omitted, see Brown (1971)

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

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

with $E\left(x_{t-1}^2\right)$ *and* $E\left(x_{t-1}^4\right)$ *given in (I.15) and (I.16) respectively.*

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

$$\frac{1}{T}\sum x_t x_{t-1} \overset{P}{\to} E\left(x_t x_{t-1}\right) = 0,$$

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

$$\frac{1}{\sqrt{T}}\sum x_t x_{t-1} \overset{D}{\to} N\left(0, E\left(x_t^2 x_{t-1}^2\right)\right).$$

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

$$\frac{1}{T}\sum_{1}^{T}\left(x_t^2 - \left(\frac{1}{T}\sum_{t=1}^{T} x_t^2\right)\right)x_{t-1}^2. \tag{I.26}$$

Considerations as above lead to the requirement that $Ex_t^4 < \infty$ *for convergence in probability, while by using Theorem I.4.1 the requirement is* $Ex_t^8 < \infty$ *for convergence in distribution. These are quite strong restrictions, and therefore to avoid such, often the autocorrelation function is given for* $|x_t|$ *instead.*

References

Bera and Higgins, 1993, ARCH Models: Properties and Estimation, Journal of Economic Surveys, 305-366

Bollerslev, T., R.Y. Chou and K.F. Kroner, 1992, ARCH modelling in Finance: A review of the theory and empirical evidence, Journal of Econometrics, 52, 5-61

Carrasco and Chen, 2002, β-mixing and Moment properties of Various GARCH, Stochastic Volatility and ACD Models, Econometric Theory

Engle, R.F., 1982, Autoregressive conditional heteroscedaticity with estimates of United Kingdom inflation, Econometrica, 50, 987-1007

Francq, C. and J.-M. Zakoïan, 2019, GARCH Models: Structure, Statistical Inference and Financial Applications, Wiley.

Franses, P.H. and D. van Dijk, 2000, Nonlinear Time Series Models in Empirical Finance, Cambridge University Press

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

Hansen, E. and A. Rahbek, 1998, Stationarity and Asymptotics of Multivariate ARCH Time Series with an Application to Robustness of Cointegration Analysis, preprint no.*12*, Department of Theoretical Statistics

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

Nelson, D., 1990, Stationarity and persistence in the GARCH(1,1) model, Econometric Theory, 6, 318-334

Pagan, A., 1996, The econometrics of financial markets, Journal of Empirical Finance, 3:15-102

Shephard, N., 1996, Statistical Aspects of ARCH and Stochastic Volatility, in "Time Series Models", ed. Cox, Hinkley and Barndorff-Nielsen, Chapman Hall, London.

Taylor, S.J., 2005, Asset Price Dynamics, Volatility and Prediction, Princeton University Press

Tjøstheim, D., 1990, Non-linear time series and markov chains, Advances in Applied Probability, 22, 587-611