Anders Rahbek
Rasmus Søndergaard Pedersen
University of Copenhagen

September 2023

Part II
General ARCH and GARCH

Similar to empirical analyses of data from macroeconomics based on autoregressive models, it is often needed in financial applications to allow for richer dynamics than the first order ARCH for example by adding further lagged values of squared returns in the conditional volatility.

We discuss in this part how the previous results can be extended and applied to such cases. It turns out that the concept of stationary, or stable, dynamics is closely related to that of the largest eigenvalue of matrices and some theory for calculus with matrices will therefore be introduced.

II.1 The ARCH(2) process

The ARCH(2) process can for $t = 2, 3, 4 \ldots$ be represented as

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

$$\sigma_t^2 = \sigma^2 + \alpha_1 x_{t-1}^2 + \alpha_2 x_{t-2}^2 \tag{II.2}$$

with initial values x_0, x_1 and where the z_t's are $i.i.d.$N(0,1). This way the distribution of x_t conditional on the information up to time $t-1$, given by the variables $(x_{t-1}, ..., x_0)$, depends on (x_{t-1}, x_{t-2}), unlike the ARCH(1).

II.1.1 Assumption I.3.1 for the ARCH(2) process

Clearly x_t does not satisfy Assumption I.3.1. However, the so-called companion satisfies Assumption I.3.1 (i) where the companion form is given by defining,

$$X_t = \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix}. \tag{II.3}$$

By this choice X_t conditional on the past values $(X_{t-1}, ..., X_0)$ depends only on $X_{t-1} = (x_{t-1}, x_{t-2})'$,

$$X_t = \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} \sigma_t z_t \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} \sqrt{\left(\sigma^2 + \alpha_1 x_{t-1}^2 + \alpha_2 x_{t-2}^2\right)} z_t \\ x_{t-1} \end{pmatrix}.$$

However, if we condition on X_{t-1} we condition on x_{t-1} in particular such that the conditional density of X_t given X_{t-1} is not satisfying Assumption I.3.1 (ii) as it is singular. We can see this directly by using the formula for conditional densities in (I.8), $f(x,y) = f(x|y)\,f(y)$, from which we would get

$$\begin{aligned} f(X_t|X_{t-1}) &= f\left((x_t, x_{t-1}) \mid (x_{t-1}, x_{t-2})\right) = f(x_t|x_{t-1}, x_{t-2})\, f(x_{t-1}|x_{t-1}, x_{t-2}) \\ &= f(x_t|x_{t-1}, x_{t-2})\, f(x_{t-1}|x_{t-1}), \end{aligned}$$

where $f(x_{t-1}|x_{t-1}, x_{t-2}) = f(x_{t-1}|x_{t-1})$, since x_{t-1} is fixed, and hence singular. That is, while indeed the first term on the right hand side $f(x_t|x_{t-1}, x_{t-2})$ is a continuous Gaussian density by definition of the ARCH(2),

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

the second term $f(x_{t-1}|x_{t-1})$ is singular.

By definition, the equation for x_{t-1} is given by,

$$x_{t-1} = \sigma_{t-1} z_{t-1} = \sqrt{\left(\sigma^2 + \alpha_1 x_{t-2}^2 + \alpha_2 x_{t-3}^2\right)}\, z_{t-1},$$

such that if we conditioned on x_{t-2} and x_{t-3} instead, that is $X_{t-2} = (x_{t-2}, x_{t-3})'$, the density of x_{t-1} given (x_{t-2}, x_{t-3}) would be well-defined and Gaussian.

Hence, it appears that while X_t conditional on X_{t-1} is not nice in the sense that it has a singular density, then X_t conditional on X_{t-2} would be satisfying a continuity condition.

To derive the density for X_t conditional on X_{t-2} we can use the formula (I.8) for conditional densities as follows,

$$\begin{aligned} f(X_t|X_{t-2}) &\overset{\text{(by definition)}}{=} f(x_t, x_{t-1}|x_{t-2}, x_{t-3}) \\ &\overset{\text{(by (I.8))}}{=} \frac{f(x_t, x_{t-1}, x_{t-2}, x_{t-3})}{f(x_{t-2}, x_{t-3})} \\ &\overset{\text{(multiply and divide)}}{=} \left(\frac{f(x_t, x_{t-1}, x_{t-2}, x_{t-3})}{f(x_{t-1}, x_{t-2}, x_{t-3})}\right)\left(\frac{f(x_{t-1}, x_{t-2}, x_{t-3})}{f(x_{t-2}, x_{t-3})}\right) \\ &\overset{\text{(by (I.8))}}{=} f(x_t|x_{t-1}, x_{t-2})\, f(x_{t-1}|x_{t-2}, x_{t-3}). \end{aligned} \tag{II.4}$$

As noted, $f(x_t|x_{t-1}, x_{t-2})$ and $f(x_{t-1}|x_{t-2}, x_{t-3})$ are Gaussian densities, and hence $f(X_t|X_{t-2})$ would satisfy Assumption I.3.1 (ii) as it is the product of two continuous densities.

To allow for the general case of more lags, k say, in the conditional variance and mean equations we can modify Assumption I.3.1 as follows:

Assumption II.1.1 *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-k}*, for some* $k \geq 1$*, has a positive conditional density* $f(X_t|X_{t-k})$*,* $f(X_t|X_{t-k}) > 0$*, which is continuous.*

And we note that Theorem I.3.2 indeed holds with the drift criterion, that is Assumption I.3.2, satisfied and under Assumption II.1.1 instead of Assumption I.3.1.

Next we turn to the drift criterion for the ARCH(2) process.

II.1.2 The drift criterion for the ARCH(2)

The theory here highlights the reason why Assumption I.3.2 allows for a general lag m when computing $E(\delta(X_t)|X_{t-m})$, where we in the previous examples used $m = 1$. It is closely related to our previous discussion on the modifed Assumption II.1.1 which was needed for the ARCH(2) process, and of course, ARCH(k) processes in general.

We start by considering the AR(2) process, and use this to introduce some few results for matrices. Recall first the concept of an eigenvalue:

Definition II.1.1 *With* A *a* $p \times p$ *matrix,*

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pp} \end{pmatrix},$$

where $a_{ij} \in \mathbb{R}$*, then the eigenvalues*[1] *of* A*,* λ_i*,* $i = 1, .., p$*, are found by solving,*

$$\det(\lambda I_p - A) = 0. \tag{II.5}$$

[1] Note that solving the equation in (II.5) amounts to find the roots of a p^{th} order polynomial, and that the roots λ may be real valued or complex valued, that is $\lambda \in \mathbb{R}$ or $\lambda \in \mathbb{C}$. If a root is complex valued one can write it as, $\lambda = a + \mathbf{i}b$, where $a, b \in \mathbb{R}$ while $\mathbf{i}^2 = -1$. An implication is that the absolute value of λ is simple to compute by,

$$|\lambda| = \sqrt{a^2 + b^2}.$$

We denote by γ_A the absolute value of the largest eigenvalue of A, that is

$$\gamma_A = \max_{i=1,\ldots,p} |\lambda_i|.$$

with the largest absolute value, $|\lambda|$, by γ_A.

Example II.1.1 *With,*

$$A = \begin{pmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{pmatrix}, \; \lambda I_2 - A = \begin{pmatrix} \lambda - \rho_1 & -\rho_2 \\ -1 & \lambda \end{pmatrix},$$

and we find

$$\det(\lambda I_2 - A) = \lambda(\lambda - \rho_1) - \rho_2 = \lambda^2 - \lambda\rho_1 - \rho_2. \tag{II.6}$$

Hence the two eigenvalues λ_1 and λ_2 of A are the two roots in this second order polynomial.

The main reason that we are interested in γ_A is that it plays a cruical role in linear as well as nonlinear time series when determining whether a time series with general lag structure is stationary.

Example II.1.2 *Consider the AR(2) process,*

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \varepsilon_t$$

with ε_t i.i.d.N(0,σ^2). Recall from the well-known theory of VAR processes, that x_t has a stationary solution if the roots in the characteristic polynomial for x_t, written here as,

$$\lambda^2 - \lambda\rho_1 - \rho_2 = 0, \tag{II.7}$$

are smaller than one in absolute value.

Next, similar to the ARCH(2), consider the companion form of the AR(2) process with $X_t = (x_t, x_{t-1})'$, such that

$$X_t = \begin{pmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{pmatrix} X_{t-1} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} = AX_{t-1} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}. \tag{II.8}$$

From Example II.1.1, we see that it is the size of the eigenvalues of A, γ_A that defines the dynamics of X_t and hence of x_t. That is the roots of the characteristic polynomial, see (II.7) and (II.6).

Analogous to the ARCH(2) case, we also see that X_t satisfies Assumption II.1.1, see (II.4) where $f(X_t|X_{t-2}) = f(x_t|x_{t-1}, x_{t-2}) f(x_{t-1}|x_{t-2}, x_{t-3})$ with,

$$f(x_t|x_{t-1}, x_{t-2}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_t - \rho_1 x_{t-1} - \rho_2 x_{t-2})^2\right).$$

Thus we have seen that the ARCH(2) and AR(2) processes satisfy Assumption II.1.1 and that the largest value of the eigenvalues of the companion form matrix for the AR(2) process determines the stationarity properties.

We next proceed to illustrate that the conclusion can be reached by using the drift criterion in Assumption I.3.2, and that we may therefore also discuss mixing, moments and stationarity at the same time.

We shall need a final result relating eigenvalues of a matrix A with the size, or norm $|A|$, of the matrix.

Definition II.1.2 *Define a matrix norm $|A|$ by*

$$|A| = \sqrt{\gamma_{A'A}}.$$

With this definition of matrix norm we have two key implications:

Lemma II.1.1 *With A a $p \times p$ dimensional matrix, then*

$$|A^m|^{1/m} \to \gamma_A,$$

as $m \to \infty$ and where γ_A is defined in Definition II.1.1 and $|A|$ is defined in Definition II.1.2. Moreover, for any vector x of dimension p,

$$|Ax| \leq |A||x|,$$

where for any vector $x = (x_1, x_2, ..., x_p)'$, $|x| = \sqrt{x_1^2 + ... + x_p^2}$.

Example II.1.3 *Consider A, given by*

$$A = \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix},$$

then as A has two eigenvalues at $\lambda = 0$, the maximal $\gamma_A = 0$. This is independent of the magnitude of the entry a, such that indeed a could be huge, while still having small γ_A. Computing, $\gamma_{A'A}$ with

$$A'A = \begin{pmatrix} 0 & 0 \\ a & 0 \end{pmatrix} \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & a^2 \end{pmatrix},$$

we find $\gamma_{A'A} = a^2$ is large if a is (in absolute value). Note that $|a| = \sqrt{a^2} = \sqrt{\gamma_{A'A}}$.

Example II.1.4 *With A as in Example II.1.3, we have*

$$A = \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}, \ A^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = A^3 = ... = A^m.$$

Thus while $|A| = \sqrt{a^2}$, after one iteration or multiplication, we have $|A^2| = \sqrt{\gamma_0} = 0 = \gamma_A$.

With $x = (x_1, x_2)'$,

$$Ax = \begin{pmatrix} ax_2 \\ 0 \end{pmatrix}, \ |Ax|^2 = a^2 x_2^2 = |A|^2 x_2^2 \leq |A|^2 \left(x_2^2 + x_1^2 \right) = |A|^2 |x|^2.$$

Example II.1.5 *With A diagonal, we get*

$$A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad \text{and } A^m = \begin{pmatrix} a^m & 0 \\ 0 & b^m \end{pmatrix}.$$

Suppose that $a > b \geq 0$ in which case $\gamma_A = a$. Next, we find

$$A'A = \begin{pmatrix} a^2 & 0 \\ 0 & b^2 \end{pmatrix} \text{ and } A^{m\prime} A^m = \begin{pmatrix} a^{2m} & 0 \\ 0 & b^{2m} \end{pmatrix}.$$

Hence $|A| = \sqrt{\gamma_{A'A}} = \sqrt{a^2} = a$, $|A^m| = a^m$ and hence for any m, $|A^m|^{1/m} = (a^m)^{1/m} = a = \gamma_A$.

Finally, observe that with $x = (x_1, x_2)'$, $Ax = (ax, bx)'$ such that

$$|Ax|^2 = a^2 x_1^2 + b^2 x_2^2 \leq a^2 \left(x_1^2 + x_2^2 \right) = |A|^2 |x|^2.$$

We are now in position to use the drift criterion on the AR(2) process.

Example II.1.6 *Consider the AR(2) process in Example II.1.2 written on companion form,*

$$X_t = \begin{pmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{pmatrix} X_{t-1} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} = AX_{t-1} + e_t. \tag{II.9}$$

A natural drift function $\delta(Y_t)$ is given by

$$\begin{aligned} \delta(X_t) &= \delta(x_t, x_{t-1}) = 1 + x_{t-1}^2 + x_{t-2}^2 \\ &= 1 + X_t' X_t = 1 + |X_t|^2. \end{aligned}$$

We can then compute $E\left(\delta\left(X_t\right)|X_{t-1}\right)$ as follows,

$$\begin{aligned} E(\delta(X_t)|X_{t-1}) &= 1 + E\left((AX_{t-1}+e_t)'\left(AX_{t-1}+e_t\right)|X_{t-1}\right) \\ &= 1 + X_{t-1}'A'AX_{t-1} + E\left(e_t'e_t\right) \\ &= 1 + X_{t-1}'A'AX_{t-1} + \sigma^2 \\ &= |AX_{t-1}|^2 + 1 + \sigma^2 \\ &\leq |A|^2|X_{t-1}|^2 + 1 + \sigma^2 \\ &= |A|^2\delta\left(X_{t-1}\right) + \left(1+\sigma^2-|A|^2\right), \end{aligned}$$

where we have used Lemma II.1.1 for the inequality. Hence it appears that choosing $|A|^2 < 1$ is an obvious choice for the drift criterion to be satistied. However, $|A|^2 < 1$ restricts the AR(2) process in a meaningless way. In particular, with $\rho_1 = \rho_2 = 0$, such that $x_t = \varepsilon_t$ and thus "as stationary as possible", $|A|^2 = 1$ which violates the condition.

We thus conclude that the condition is too strong (actually meaningless). What should be considered instead is the recursion of the dynamic system for X_t. More precisely, as in the univariate AR(1) case, we can do simple recursion in (II.9),

$$\begin{aligned} X_t &= AX_{t-1} + e_t = A^2X_{t-2} + Ae_{t-1} + e_t = \dots \\ &= A^mX_{t-m} + \sum_{i=0}^{m-1} A^i e_{t-i}. \end{aligned}$$

And we find,

$$E\left(\delta(X_t)|X_{t-m}\right) = 1 + X_{t-m}'A^{m\prime}A^mX_{t-1} + \tilde{\sigma}^2,$$

with $\tilde{\sigma}^2$ proportional to σ^2. As before, we can use Lemma II.1.1,

$$X_{t-m}'A^{m\prime}A^mX_{t-m} = |A^mX_{t-m}|^2 \leq |A^m|^2|X_{t-m}|^2,$$

such that the drift criterion is satisfied if $|A^m| < 1$. Now we can apply Lemma II.1.1 from which it holds that,

$$|A^m|^{1/m} \to \gamma_A.$$

We can therefore conclude that if $\gamma_A < 1$, then also $|A^m| < 1$ for some m large enough. And the condition $\gamma_A < 1$ is exactly the well-known condition for stationarity of an AR(2) process, see Example II.1.2.

II.1.2.1 The ARCH(2) process

Return to the ARCH(2) process as defined by (II.1),

$$x_t = \sigma_t z_t, \quad \sigma_t^2 = \sigma^2 + \alpha_1 x_{t-1}^2 + \alpha_2 x_{t-2}^2,$$

which has companion form for $X_t = (x_t, x_{t-1})'$ in (II.9).).

Lemma II.1.2 *With the ARCH(2) process given by (II.1), then $X_t = (x_t, x_{t-1})'$ satisfies Theorem I.3.2 and has finite second order moments if either of the three equivalent conditions hold for $\alpha_1, \alpha_2 \geq 0$:*

(i) $\quad \alpha_1 + \alpha_2 < 1$

(ii) $\quad |\lambda^2 - \alpha_1 \lambda - \alpha_2| = 0 \Rightarrow |\lambda| < 1$

(iii) $\quad \gamma_A < 1$ *with* $A = \begin{pmatrix} \alpha_1 & 1 \\ \alpha_2 & 0 \end{pmatrix}$

In particular, under either of (i), (ii) and (iii), x_0 and x_1 can be given an initial distribution such that x_t (and X_t) is stationary, weakly mixing and $Ex_t^2 < \infty$.

Proof of Lemma II.1.2:
Start by computing $E(\delta(X_t)|X_{t-m}) = E(\delta(X_t)|x_{t-m}, x_{t-m-1})$, for $m = 1$,

$$\begin{aligned} E(\delta(X_t)|X_{t-1}) &= E(1 + X_t' X_t | X_{t-1}) \\ &= 1 + E\left(\left(x_t^2 + x_{t-1}^2\right)|x_{t-1}, x_{t-2}\right) = E\left(\left(\sigma_t^2 z_t^2 + x_{t-1}^2\right)|x_{t-1}, x_{t-2}\right) \\ &= 1 + \sigma_t^2 + x_{t-1}^2 \\ &= 1 + \sigma^2 + (\alpha_1 + 1) x_{t-1}^2 + \alpha_2 x_{t-2}^2. \end{aligned} \tag{II.10}$$

Next, using Lemma I.2.1 for conditional expectations, and (II.10), we see that for $m = 2$ we get,

$$\begin{aligned} E(\delta(X_t)|X_{t-2}) &= E(E(\delta(X_t)|X_{t-1}, X_{t-2})|X_{t-2}) \\ &= E(E(\delta(X_t)|X_{t-1})|X_{t-2}) \\ &= E\left(1 + \sigma^2 + (\alpha_1 + 1) x_{t-1}^2 + \alpha_2 x_{t-2}^2 | x_{t-2}, x_{t-3}\right) \\ &= 1 + \sigma^2 + (\alpha_1 + 1) \sigma_{t-1}^2 + \alpha_2 x_{t-2}^2, \end{aligned}$$

such that

$$E(\delta(X_t)|X_{t-2}) = 1 + \sigma^2 + (\alpha_1 + 1)\sigma^2 + ((\alpha_1 + 1)\alpha_1 + \alpha_2) x_{t-2}^2 + (\alpha_1 + 1)\alpha_2 x_{t-3}^2. \tag{II.11}$$

And, likewise for $m = 3$[2],

$$
\begin{aligned}
E\left(\delta\left(X_t\right)|X_{t-3}\right) &= E\left(E\left(\delta\left(X_t\right)|X_{t-2}\right)|X_{t-3}\right) \\
&= 1 + \sigma^2 + \left(\alpha_1 + 1\right)\sigma^2 + \left(\left(\alpha_1 + 1\right)\alpha_1 + \alpha_2\right)\sigma^2 \\
&+ \left(\left(\left(\alpha_1 + 1\right)\alpha_1 + \alpha_2\right)\alpha_1 + \left(\alpha_1 + 1\right)\alpha_2\right)x_{t-3}^2 + \left(\left(\alpha_1 + 1\right)\alpha_1 + \alpha_2\right)\alpha_2 x_{t-4}^2
\end{aligned}
\tag{II.13}
$$

To define the coefficients appearing in the conditional expectations above we define,

$$\beta_0 = 1,\ \beta_1 = \alpha_2, \quad \beta_2 = \left(1 + \alpha_1\right)$$

and next for $m = 2, 3, ...$

$$\beta_{2m} = \alpha_1 \beta_{2m-2} + \beta_{2m-3}, \quad \text{and} \quad \beta_{2m-1} = \alpha_2 \beta_{2m-2}. \tag{II.14}$$

With this definition, for example for $m = 3$ in (II.13) we find immediately,

$$E\left(\delta\left(X_t\right)|X_{t-3}\right) = 1 + \beta_0\sigma^2 + \beta_2\sigma^2 + \beta_4\sigma^2 + \beta_6 x_{t-3}^2 + \beta_5 x_{t-4}^2,$$

and for general m,

$$E\left(\delta\left(X_t\right)|X_{t-m}\right) = E\left(\delta\left(X_t\right)|x_{t-m}, x_{t-m-1}\right) = \beta_{2m} x_{t-m}^2 + \beta_{2m-1} x_{t-m-1}^2 + \tilde{\sigma}_m^2 \tag{II.15}$$

where $\tilde{\sigma}_m^2 = \sum_{i=0}^{m-1} \beta_{2i}\sigma^2$.

We can write the right hand side of (II.15) as follows,

$$
\begin{aligned}
E\left(\delta\left(X_t\right)|X_{t-m}\right) &= X_{t-m}' \begin{pmatrix} \beta_{2m} & 0 \\ 0 & \beta_{2m-1} \end{pmatrix} X_{t-m} + \tilde{\sigma}_m^2 \\
&= |BX_{t-m}|^2 + \tilde{\sigma}_m^2 \\
&\leq |B|^2 |X_{t-m}|^2 + \tilde{\sigma}_m^2,
\end{aligned}
$$

[2] Also, for $m = 4$,

$$
\begin{aligned}
E\left(\delta\left(X_t\right)|X_{t-4}\right) &= E\left(E\left(\delta\left(X_t\right)|X_{t-3}\right)|X_{t-4}\right) \\
&= 1 + \sigma^2 + \left(\alpha_1 + 1\right)\sigma^2 + \left(\left(\alpha_1 + 1\right)\alpha_1 + \alpha_2\right)\sigma^2 + \\
&\left(\left(\left(\alpha_1 + 1\right)\alpha_1 + \alpha_2\right)\alpha_1 + \left(\alpha_1 + 1\right)\alpha_2\right) E\left(x_{t-3}^2 | x_{t-4}, x_{t-5}\right) + \\
&\left(\left(\alpha_1 + 1\right)\alpha_1 + \alpha_2\right)\alpha_2 x_{t-4}^2 \\
&= 1 + \sigma^2 + \left(\alpha_1 + 1\right)\sigma^2 + \left(\left(\alpha_1 + 1\right)\alpha_1 + \alpha_2\right)\sigma^2 + \\
&\left(\left(\left(\alpha_1 + 1\right)\alpha_1 + \alpha_2\right)\alpha_1 + \left(\alpha_1 + 1\right)\alpha_2\right)\sigma^2 + \\
&\left(\left(\left(\left(\alpha_1 + 1\right)\alpha_1 + \alpha_2\right)\alpha_1 + \left(\alpha_1 + 1\right)\alpha_2\right)\alpha_1 + \left(\left(\alpha_1 + 1\right)\alpha_1 + \alpha_2\right)\alpha_2\right) x_{t-4}^2 \\
&\left(\left(\left(\alpha_1 + 1\right)\alpha_1 + \alpha_2\right)\alpha_1 + \left(\alpha_1 + 1\right)\alpha_2\right)\alpha_2 x_{t-5}^2
\end{aligned}
\tag{II.12}
$$

where

$$B = \left(\begin{array}{cc} \sqrt{\beta_{2m}} & 0 \\ 0 & \sqrt{\beta_{2m-1}} \end{array} \right).$$

Thus as in Example II.1.6 we need B to have eigenvalues smaller than 1. But the eigenvalues of B are simply $\lambda_1 = \sqrt{\beta_{2m}}$ and $\lambda_2 = \sqrt{\beta_{2m-1}}$, so what we need is $\beta_{2m} < 1$ and $\beta_{2m-1} < 1$.

To see that this holds, use the definition of the coefficients β_{2m} and β_{2m-1} in (II.14), and write these as,

$$v_m = \left(\begin{array}{c} \beta_{2m} \\ \beta_{2m-1} \end{array} \right) = A \left(\begin{array}{c} \beta_{2m-2} \\ \beta_{2m-3} \end{array} \right) = Av_{m-1}, \text{ with } A = \left(\begin{array}{cc} \alpha_1 & 1 \\ \alpha_2 & 0 \end{array} \right).$$

Thus with $\gamma_A < 1$, then by Lemma II.1.1, $|A^m| < 1$ for m large and hence as

$$v_m = A^m e, \quad \text{with } e = (1,1)'$$

we get

$$|v_m|^2 = v_m' v_m = e' A^{m\prime} A^m e = |A^m e|^2 \leq |A^m|^2 |e|^2 = 2|A^m|^2 < 1 \text{ for } m \text{ large.}$$

So the v_m elements, that is the diagonal elements of B, are indeed smaller than 1 in absolute value as desired.

II.2 ARCH(k)

The ARCH(k) process can for $t = k, k+1, k+2 \ldots$ be represented as

$$x_t = \sigma_t z_t \tag{II.16}$$
$$\sigma_t^2 = \sigma^2 + \alpha_1 x_{t-1}^2 + ... + \alpha_k x_{t-k}^2 \tag{II.17}$$

with initial values $x_0, x_1, ..., x_{k-1}$ and where the z_t's are $i.i.d.$N(0,1). As in the ARCH(2) case, we can use the drift criterion to establish that the ARCH(k) process with $X_t = (x_t, x_{t-1}, ..., x_{t-k-1})'$ satisfies Theorem I.3.2 and has finite second order moments if, $\alpha_i \geq 0$ and

$$\left| \lambda^k - \alpha_1 \lambda^{k-1} - ... - \alpha_k \right| = 0 \Rightarrow |\lambda| < 1. \tag{II.18}$$

This is equivalent to the condition that,

$$\alpha_1 + \alpha_2 + ... + \alpha_k < 1.$$

II.2.1 Multivariate ARCH(k)

Consider one natural extension of the univariate ARCH process to the $p-$dimensional process as given by

$$X_t = \Omega_t^{1/2} Z_t, \qquad \text{(II.19)}$$
$$\Omega_t = \Omega + \sum_{i=1}^{k} A_i X_{t-i} X_{t-i}' A_i',$$

where $\Omega > 0$, (A_i) are any $p \times p$ matrices and Z_t $i.i.d.N_p\left(0, I_p\right)$. This is a simple example of the so-called BEKK process in Engle and Kroner (1995). In the scalar-ARCH the A_i matrices are replaced by scalars $\alpha_i (= A_i^2)$ which is sometimes applied in high-dimensional, or "vast", systems.

Other extensions include that at each lag $i = 1, ..., k$ further loadings may be added, such that with $s \leq p$, one may use the general ARCH,

$$X_t = \Omega_t^{1/2} Z_t, \qquad \text{(II.20)}$$
$$\Omega_t = \Omega + \sum_{i=1}^{k} \sum_{j=1}^{s} A_{ij} X_{t-i} X_{t-i}' A_{ij}',$$

Here the summation over j for fixed i, allows for a more rich feed-back mechanism from X_{t-i} which in the univariate case vanishes.

Mimicking the just given calculations for the univariate ARCH(2) it follows that if

$$\left| I_{p^2} \lambda^k - \sum_{j=1}^{s} \left(A_{1j} \otimes A_{1j}\right) \lambda^{k-1} - ... \sum_{j=1}^{s} \left(A_{kj} \otimes A_{kj}\right) \right| = 0 \Rightarrow |\lambda| < 1,$$

holds, then X_t satisfies Theorem I.3.2 with $\delta\left(X_t\right) = 1 + |X_t|^2$, has second order moments and can be given an initial distribution such that it becomes stationary.

Note that with A a $p \times p$ matrix,

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pp} \end{pmatrix},$$

then $A \otimes A$ is the $\left(p^2 \times p^2\right)-$dimensional matrix defined by,

$$A \otimes A = \begin{pmatrix} a_{11}A & \cdots & a_{1p}A \\ \vdots & & \vdots \\ a_{p1}A & \cdots & a_{pp}A \end{pmatrix}.$$

II.3 GARCH(1,1)

The generalized ARCH process, the GARCH(1,1) is probably the most used of all ARCH processes. The GARCH(1,1), or simply the GARCH, process for $t = 1, 2, ...$ can be represented as

$$x_t = \sigma_t z_t \tag{II.21}$$

$$\sigma_t^2 = \sigma^2 + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{II.22}$$

with initial values x_0 and σ_0^2, and where the z_t's are $i.i.d.$N(0,1). Moreover, $\sigma^2 > 0$, $\alpha \geq 0$ and $\beta \geq 0$.

Loosely speaking, the GARCH process can be motivated by a few recursions:

$$\begin{aligned}
\sigma_t^2 &= \sigma^2 + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2 \\
&= \sigma^2 + \alpha x_{t-1}^2 + \beta \left(\sigma^2 + \alpha x_{t-2}^2 + \beta \sigma_{t-2}^2\right) \\
&= (1+\beta)\,\sigma^2 + \alpha x_{t-1}^2 + \beta \alpha x_{t-2}^2 + \beta^2 \sigma_{t-3}^2 \\
&= \left(1+\beta+\beta^2\right)\sigma^2 + \alpha x_{t-1}^2 + \beta \alpha x_{t-2}^2 + \beta^2 \alpha x_{t-3}^2 + \beta^3 \sigma_{t-3}^3 \\
&= ... \\
&= \sum_{i=0}^{t-1} \beta^i \sigma^2 + \sum_{i=0}^{t-1} \alpha \beta^i x_{t-1-i}^2 + \beta^t \sigma_0^2
\end{aligned}$$

Hence, it may be viewed as a way in just two parameters, α and β, to allow for an increasing (in t) number of lagged x_t^2 to enter the ARCH specification.

Now x_t conditional on the past information $(x_{t-1}, ..., x_0)$ and σ_0^2 is not a function of x_{t-1} but a function of the entire past $(x_{t-1}, x_{t-2}, ..., x_0)$ and also σ_0^2. So it is not a Markov chain. However, by setting for example $X_t = (x_t, \sigma_t)'$ then indeed X_t given past X_t is a function of X_{t-1}, and thus satisfying Assumption II.1.1 (i). Unfortunately defining X_t this way, Assumption II.1.1 is not satisfied. Instead the following approach can be applied, see Carrasco and Chen (2002) for example. Rewrite σ_t^2 as follows,

$$\begin{aligned}
\sigma_t^2 &= \sigma^2 + \alpha \sigma_{t-1}^2 z_{t-1}^2 + \beta \sigma_{t-1}^2 \\
&= \sigma^2 + \left(\alpha z_{t-1}^2 + \beta\right) \sigma_{t-1}^2.
\end{aligned}$$

Then we can apply the drift criterion to σ_t and next use, $x_t = \sqrt{\sigma_t^2} z_t$ to conclude that x_t inherits the properties of σ_t^2. More precisely, cf. Carrasco and Chen (2002, Proposition 5) (Meitz and Saikkonen, 2008, discuss modifications of Proposition 5 in Carrasco and Chen, 2002), if α and β are such that Theorem I.3.2 holds for a drift function $\delta\left(\sigma_t^2\right)$, then σ_t^2 can be given an initial distribution such that σ_t^2 and hence x_t are stationary and weakly mixing. Moreover, as $E\delta\left(\sigma_t^2\right) < \infty$, then if $\delta\left(\sigma_t^2\right) = 1 + \left(\sigma_t^2\right)^s$, $s > 0$, we also have $Ex_t^{2s} < \infty$.

Consider the simple example of $\delta\left(\sigma_t^2\right) = 1 + \sigma_t^2$. Then

$$\begin{aligned} E\left(\delta\left(\sigma_t^2\right)|\sigma_{t-1}^2\right) &= \sigma^2 + \alpha\sigma_{t-1}^2 + \beta\sigma_{t-1}^2 \\ &= (\alpha+\beta)\,\delta\left(\sigma_{t-1}^2\right) + \left(\sigma^2 - (\alpha+\beta)\right). \end{aligned}$$

Hence we immediately conclude that if $(\alpha+\beta) < 1$, then $Ex_t^2 < \infty$ and x_t and σ_t^2 are weakly mixing.

More generally, see also Nelson (1990, Theorem 3), we have the following similar to the ARCH(1) process:

GARCH(1,1) process x_t defined in (II.21):	
$x_t = \sigma_t z_t,\ \ \sigma_t = \sigma^2 + \alpha x_{t-1}^2 + \beta\sigma_{t-1}^2$ and z_t $i.i.d.\mathrm{N}(0,1)$.	
Stationary for $E\log\left(\alpha z_t^2 + \beta\right) < 0$	
Finite moments:	
$E\log\left(\alpha z_t^2 + \beta\right) < 0$	no moments (fractional)
$E\left(\alpha z_t^2 + \beta\right)^{1/2} < 1$	$E\lvert x_t\rvert < \infty$
$E\left(\alpha z_t^2 + \beta\right) < 1$ or $(\alpha+\beta) < 1$	$Ex_t^2 < \infty$
$E\left(\alpha z_t^2 + \beta\right)^2$ or $\beta^2 + 3\alpha^2 + 2\alpha\beta < 1$	$Ex_t^4 < \infty$

In general, the condition is that $E\left(\sigma_t^{2k}\right) < \infty$ if and only if $E\left(\alpha z_t^2 + \beta\right)^k < 1$, see Nelson (1990) and Carrasco and Chen (2002). Explicit expressions for $E\log\left(\alpha z_t^2 + \beta\right)$ and $E\left(\alpha z_t^2 + \beta\right)^k$ in terms of α, β and so-called hypergeometric functions are given in Nelson (1990). Note also that the regions change if z_t is assumed to be $i.i.d.$ t_v distributed, for example.

Lastly, an example of a multivariate GARCH is the simple BEKK-GARCH(1,1) process given by,

$$\Omega_t = \Omega + AX_{t-1}X_{t-1}'A' + B\Omega_{t-1}B',$$

with A and B $(p \times p)$-dimensional matrices.

References

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

Engle and Kroner,(1995), Multivariate Simultaneous Generalized ARCH,. Econometric Theory

Meitz and Saikkonen, (2008), Ergodicity, Mixing, and Existence of Moments of a Class of Markov Models with Applications to GARCH and ACD Models, Econometric Theory.

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