Chapter 3

Heavy-tailed time series

Linear time series with finite variance and, in particular, normally distributed ones played a crucial role in the development of time series analysis. This also explains the important role played by autocorrelations in standard time series analysis. It is not completely surprising that the asymptotic theory of the sample autocorrelations breaks down in the presence of non-linear heavy-tailed models like the simple bilinear model suggested by Davis and Resnick [19]. In Section 3.1 we recall some facts about linear time series in the light- and heavy-tailed cases and introduce some non-linear models as well. In Section 3.2 special attention is given to the study of stochastic recurrence equations. We present some of the most popular non-linear time series models that originate from these equations and are used in the econometrics literature nowadays. We provide some basic results for their asymptotic analysis in Chapter 4 by showing that they give rise to regularly varying strongly mixing stationary processes.

3.1 Linear and non-linear time series models

3.1.1 $L_2$ time series analysis

Although this chapter is dedicated to the asymptotic analysis of certain heavy-tailed time series models, we would like to compare our results with those from standard time series analysis for light-tailed series. For this comparison we will recall some notions and results of this well established theory. The main reference of this section is Brockwell and Davis [12]. We may call a stationary process $(X_n)$ light-tailed if $X_1$ has a finite second moment. Therefore we may refer to this theory as $L_2$ time series analysis. A special role in this respect is reserved for mean-zero Gaussian stationary processes, since their distribution is completely determined by the second moments, i.e. variances and covariances.

Heavy-tailed processes can be defined in various ways. There is no general agreement in the literature about the meaning of the term. We will think of
a heavy-tailed time series as a stationary sequence of regularly varying random variables with index \( \alpha < 2 \). Notice that, according to our definitions, there is a whole class of processes which are neither light- nor heavy-tailed.

From the theory of Hilbert spaces we know how to give the best predictors of \( X_n \) in the mean-square distance. For a given closed subspace \( \mathcal{L} \) of \( L_2(\Omega, \mathcal{F}, P) \) the best predictor of \( X_n \) in \( \mathcal{L} \) is simply the projection of \( X_n \) onto \( \mathcal{L} \). An important example is the closed subspace \( \mathcal{L} \) spanned by all the functions of the observables. The scalar product \((X, Y)\) here is given by \( E(XY) \).

This projection can be given in terms of conditional expectations, which, in practice, can be very hard to assess. Therefore, one frequently restricts attention to linear functions of the observables. The best linear one-step predictor of \( X_{n+1} \), given the past and present \((X_t)_{t \leq n}\), is the projection of \( X_{n+1} \) onto the subspace \( \mathcal{L}_n = \text{span}\{X_i : -\infty < i \leq n\} \) of \( L_2(\Omega, \mathcal{F}, P) \), equipped with the inner product \((X, Y) = E(XY)\). We denote this one-step linear predictor by \( \pi_{\mathcal{L}_n}X_{n+1} \). In words, \( \mathcal{L}_n \) is the closure of the linear subspace spanned by the random variables \( X_i, -\infty < i \leq n \). Since \((X_n)\) is stationary, we may define the one-step prediction error as

\[
\sigma^2 = E|X_{n+1} - \pi_{\mathcal{L}_n}X_{n+1}|^2.
\]

We say that a process \((X_n)\) is deterministic if \( \sigma^2 = 0 \). Hence, for a deterministic process, \( X_{n+1} \) is completely predictable given its past and present. An important result of \( L_2 \) time series analysis is the Wold decomposition theorem, (see Brockwell and Davis [12], Theorem 5.7.1). It states that \( X_n \) can be expressed as the sum of a linear process and a deterministic process, i.e.

\[
X_n = \sum_{j=0}^{\infty} \psi_j Z_{n-j} + V_n.
\]  

(3.1)

Here \((Z_n)\) is a white noise process, i.e. an uncorrelated sequence of random variables with zero expectation and fixed variance. The process \((V_n)\) is deterministic and uncorrelated with \((Z_n)\). The constants \(\psi_n\) satisfy \(\psi_0 = 1\) and \(\sum_{j=0}^{\infty} \psi_j^2 < \infty\). Moreover the decomposition (3.1) is unique.

The Wold decomposition theorem has some interesting consequences for the \( L_2 \) theory of time series. It says that any purely non-deterministic process \((X_n)\) is linear, i.e. it is an infinite order moving average

\[
X_n = \sum_{j=0}^{\infty} \psi_j Z_{n-j},
\]  

(3.2)

based on an appropriate white noise sequence \((Z_n)\) and a sequence \((\psi_n)\) such that \(\sum_{j=0}^{\infty} \psi_j^2 < \infty\). The question arises as to whether linear processes play such an important role in the heavy-tailed case as well. We will argue later, that the answer to this question is: no! Another question is if there is a class of processes which
may play a similar role in the heavy-tailed case. To the best of our knowledge, the
answer is not known.

The Wold decomposition gives a partial reason why standard time series analysis
focuses on linear processes. In particular, the widely used class of causal autoregres-
sive moving average or ARMA processes consists of purely non-deterministic
processes. We say that a stationary process \( (X_n) \) is an ARMA\((p,q)\) process if it satisfies

\[
X_n - \phi_1 X_{n-1} - \cdots - \phi_p X_{n-p} = Z_n + \theta_1 Z_{n-1} + \cdots + \theta_q Z_{n-q},
\]

for some real constants \( \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q \) and a white noise sequence \( (Z_n) \).

For linear processes in general, and ARMA processes in particular, the estimation
of their parameters and the determination of the best linear predictors are
usually based on appropriate scalar products, i.e. autocovariances and autocor-
relations, defined as functions on \( Z \) by

\[
\gamma(h) = \text{Cov}(X_0, X_h) = E(X_0 - E X_0)(X_h - E X_h),
\]

\[
\varrho(h) = \text{Corr}(X_0, X_h) = \frac{\gamma(h)}{\gamma(0)}.
\]

These functions are estimated by their sample versions. The definitions follow.

The sample autocovariances are defined as

\[
\hat{\gamma}(h) = \frac{1}{n} \sum_{j=1}^{n-h} (X_{j+h} - \bar{X}_n)(X_j - \bar{X}_n), \quad 0 \leq h < n,
\]

where \( \bar{X}_n \) is the sample mean. We shall suppress the sample mean in the case
of processes with zero or infinite mean. Also, we divide by \( n \), instead of \( n - h \) in
order to ensure that the matrix \( \{\hat{\gamma}(i-j)\}_{i,j=1}^{n} \) is non-negative definite.

The sample autocorrelations (sample ACF) are now defined as

\[
\hat{\varrho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad 0 \leq h < n.
\]

It is a standard approach in \( L_2 \) time series analysis first to establish consistency
of these estimators (which is an immediate consequence of the ergodic theorem if
\( (X_n) \) is stationary and ergodic) and then to conclude that parameter estimators
based on them are consistent as well. For instance in the case of the linear process
as given by (3.2) with an i.i.d. sequence \( (Z_n) \) it is well known that

\[
\hat{\varrho}(h) \overset{P}{\to} \varrho(h) = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+h}}{\sum_{j=0}^{\infty} \psi_j^2}.
\]  (3.3)
In the case $E|Z_i|^2 < \infty$, under some additional conditions on the sequence $(\psi_j)$ the estimators $\hat{\theta}(h)$ are asymptotically normal, see Brockwell and Davis [12], Chapter 7.

All these well-known results further allow one to give a prediction of future values and to gain some understanding of the physical mechanism underlying a given time series. Consequently, sample autocorrelations and autocovariances play an extremely important role in the case of $L_2$ time series analysis. As mentioned earlier, their deterministic counterparts completely describe the distribution of stationary Gaussian sequences, which makes those especially interesting for modelling and prediction. It is not immediately clear, however, whether sample autocorrelations and autocovariances have any meaning in the case of heavy-tailed processes, in particular for the processes with regular variation index less than 2.

### 3.1.2 Linear time series with heavy-tailed innovations

As we said earlier, for our purposes heavy-tailed distributions are regularly varying and have infinite second moment. Davis and Resnick in [14, 15] considered linear processes

$$X_n = \sum_{j=0}^{\infty} \psi_j Z_{n-j}, \quad n \in \mathbb{Z},$$

(3.4)

where $(Z_n)_{n \in \mathbb{Z}}$ is a sequence of i.i.d. regularly varying random variables with index $\alpha \in (0, 2)$, and $(\psi_n)_{n \in \mathbb{N}}$ is a sequence of constants such that $\sum j |\psi_j|^\delta < \infty$ for some $\delta < \min(1, \alpha)$. The condition of regular variation of $Z_n$ ensures that $X_n$ is regularly varying as well (see Mikosch and Samorodnitsky [43] or Davis and Resnick [14]), while the summability condition on the coefficients $\psi_j$ implies that the series in (3.4) converges almost surely and in $L_2$.

Although $(X_n)$ is a stationary sequence, we cannot speak about its autocovariances and autocorrelations. Indeed, since $\alpha < 2$, the second moment of $X_n$ is infinite. However, we can study their sample versions. Davis and Resnick [14], [15], show that, somewhat surprisingly, for any $h \in \mathbb{N},$

$$\frac{n}{a_n^2} \left( \frac{\hat{\gamma}(0), \hat{\gamma}(1), \ldots, \hat{\gamma}(h)}{\sum_{j=0}^{\infty} \psi_j^2, \sum_{j=0}^{\infty} \psi_j \psi_{j+1}, \ldots, \sum_{j=0}^{\infty} \psi_j \psi_{j+h}} \right) \to Y_{\alpha/2},$$

where the random variable $Y_{\alpha/2}$ is $\alpha/2$-stable and positive. Therefore

$$\hat{\theta}(h) \to \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+h}}{\sum_{j=0}^{\infty} \psi_j^2} =: \vartheta(h).$$

(3.5)

The sequence $(a_n)$ above is chosen to satisfy the regular variation condition:

$$n P(Z_1 \in a_n \cdot) \to \mu_1 (\cdot),$$
where $\mu_1$ is a measure on $\mathbb{R}\setminus\{0\}$, as in (2.15). Thus we have the same weak limit for $\hat{\varrho}(h)$ as in the case of finite variance, cf. (3.3). This suggests that $\varrho(h)$ can be used as some kind of measure of dependence in an infinite variance linear process, as population autocorrelation say, analogous to the autocorrelations in a finite variance process.

Since in classical $L_2$ time series analysis the consistency of the sample ACF is the basis for the estimation of parameters and forecasting, relation (3.5) suggests that the same applies in the infinite variance case. This is true indeed, see for instance Embrechts et al. [25], Chapter 7. It is also interesting that the rate of convergence of $\hat{\varrho}(h)$ towards $\varrho(h)$ compares favourably with the classical $\sqrt{n}$-rate, see Davis and Resnick [15], cf. Brockwell and Davis [12], Section 7.2.

### 3.1.3 Non-linear time series

It is not a priori clear if a similar theory for the sample autocorrelations exists in the case of non-linear heavy-tailed processes. It was shown by various authors that the linear process case is rather exceptional, we refer to Resnick [35] for an overview paper. One of the first efforts in this direction was made by Davis and Resnick [19]. They considered a bilinear process with regularly varying innovations, i.e. they assume that the sequence $(X_n)$ satisfies

$$X_n = b X_{n-1} Z_{n-1} + Z_n, \quad n \in \mathbb{Z},$$

for some constant $b$ and an i.i.d. sequence of non-negative regularly varying random variables $Z_n$ with index $\alpha$. They showed that the condition of regular variation is inherited by the process $(X_n)$, but the index is now $\alpha/2$. The sample autocorrelations of $(X_n)$ converge in distribution as before, but now to non-degenerate limit random variables. It means that the sample autocorrelation function plot is of very little value under these circumstances.

We shall show later, in Section 3.2 that this can be true even when $(Z_n)$ is a light-tailed sequence. To say it more precisely, assume that the sequence $(X_n)$ is given by

$$X_n = a X_{n-1} + b X_{n-1} Z_{n-1} + Z_n, \quad (3.6)$$

for an i.i.d. sequence $(Z_n)$ and constants $a, b \neq 0$. The constants $a$ and $b$ can be chosen such that the stationary solution to (3.6) exists and is regularly varying with index smaller than 2. This is possible even when the innovations $Z_n$ are light-tailed, e.g. normal. Such a process $(X_n)$ is called a **simple bilinear process**. It belongs to a larger class of bilinear processes, which can be understood as one possible generalization of an ARMA process. A general **bilinear process** satisfies
the following recurrence equations

\[ X_n = \sum_{j=1}^{p} \phi_j X_{n-j} + \sum_{j=1}^{q} \theta_j Z_{n-j} + \sum_{i_1=1}^{k} \sum_{i_2=1}^{l} \beta_{i_1,i_2} X_{n-i_1} Z_{n-i_2}, \quad n \in \mathbb{Z}, \quad (3.7) \]

for some constants \( \phi_j, \theta_j \) and \( \beta_{i_1,i_2} \) and an i.i.d. sequence \( (Z_n) \).

In the econometrics literature one class of non-linear processes attracted a lot of attention and research interest over the last two decades: the class of ARCH processes. Its simplest member was introduced by Engle [26] in 1982. ARCH processes became so popular that by now more than 100 extensions and modifications of this model exist. They are again defined through recurrence equations.

![Graph](image1)

**Figure 3.1.1** A time series of 1000 daily Ford stock log-returns is plotted. The graph below represents a simulated sample path of an GARCH(1,1) process with innovations \( Z_n \) coming from the \( t \)-distribution with 6 degrees of freedom. Its parameters are estimated from the time series above as \( \alpha_0 = 8.89 \cdot 10^{-7}, \alpha_1 = 0.0166 \) and \( \beta_1 = 0.973 \), see (3.8) and (3.10).

A stationary stochastic process is called an ARCH (autoregressive condition-
3.1. Linear and non-linear time series models

ally heteroscedastic) process of order \( p \) \( (p \in \mathbb{N}) \) if it satisfies

\[
X_n = \sigma_n Z_n, \quad n \in \mathbb{Z},
\]

(3.8)

where \( (Z_n) \) is a sequence of symmetric (or only mean-zero) i.i.d. random variables with finite variance (frequently one assumes that they are standard normal) and \( \sigma_n \) is the conditional standard deviation of \( X_n \) (also called stochastic volatility), given the past observations. It satisfies the equations

\[
\sigma_n^2 = \alpha_0 + \alpha_1 X_{n-1}^2 + \cdots + \alpha_p X_{n-p}^2, \quad n \in \mathbb{Z},
\]

(3.9)

for non-negative constants \( \alpha_0, \alpha_1, \ldots, \alpha_p \), with both \( \alpha_0 \) and \( \alpha_p \) different from 0.

Financial time series, such as relative returns of stock indices, share prices and foreign exchange rates, exhibit the following typical properties (usually referred to as “stylized facts”):

- exceedances of high/low thresholds appear in clusters, indicating that there is dependence in the tails.

On the other hand,

- the sample ACF of such data is negligible at all lags (with a possible exception of the first lag).

This suggests that returns come from a white noise model. The mentioned properties are well described by an ARCH process with a sufficiently large number of parameters \( \alpha_i \).

Returns have another characteristic property:

- given their stationarity, there is statistical evidence that the 4th or 5th moment of the underlying marginal distribution might not exist.

As we will explain in Section 3.2, ARCH processes have regularly varying finite-dimensional distribution, and therefore they also capture the previously mentioned erratic behaviour of real-life data.

In order to get a good fit to real-life data one usually needs a large number of parameters \( \alpha_i \). Therefore alternative models were introduced. The generalized ARCH (or GARCH) model became the most popular among them. It was introduced by Bollerslev [8] in 1986. Since then, the GARCH model has become one of the most successful econometric time series models.

The \textbf{GARCH}(p, q) model (for \( p, q \in \mathbb{N} \)) is defined by the same formula (3.8), but the stochastic volatility now satisfies

\[
\sigma_n^2 = \alpha_0 + \alpha_1 X_{n-1}^2 + \cdots + \alpha_p X_{n-p}^2 + \beta_1 \sigma_{n-1}^2 + \cdots + \beta_q \sigma_{n-q}^2.
\]

(3.10)
where all the coefficients $\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ are nonnegative and at least $\alpha_0, \alpha_p$ and $\beta_q$ are not equal to 0.

Another well-studied non-linear model is the generalized autoregressive process, which satisfies the following difference equation

$$X_n = \Phi^{(n)}_{n-1} X_{n-1} + \cdots + \Phi^{(n)}_{n-p} X_{n-p} + Z_n.$$  \hfill (3.11)

The sequence of random vectors $(\Phi^{(n)}_{n-1}, \ldots, \Phi^{(n)}_{n-p})$ is i.i.d. and independent of the i.i.d. sequence $(Z_n)$.

The non-linear processes above have one thing in common: they can be represented or embedded into a (possibly multivariate) stochastic recurrence equation. Stochastic recurrence equations have been treated in the mathematics literature for decades, so that many of their properties follow from an already established theory.

### 3.2 Stochastic recurrence equations

#### 3.2.1 Basic theory for stochastic recurrence equations

In this section we describe general results concerning stochastic recurrence equations with the intention of applying them in the particular case of the non-linear models from the previous section. We start with a simple model with motivation in economics, which is known as “perpetuity”.

Assume that you hold an investment in a particular kind of asset of value $X_{n-1}$ between the time points $n-1$ and $n$. At the end of each time period $n$ you get paid a random interest, with the rate $r_n$. Suppose further, that during the same period you had a random income (possibly negative, if gains are less than losses) of the amount $B_n$. This amount you invest in the same type of asset. Suppose finally that $((r_n, B_n))$ is an i.i.d. sequence. Then the value of your investment is represented by a Markov chain which satisfies the 1-dimensional stochastic recurrence equation

$$X_n = A_n X_{n-1} + B_n, \quad n \in \mathbb{N},$$  \hfill (3.12)

with $A_n = 1 + r_n$.

It is interesting to ask under which constraints on the distribution of $(A_1, B_1)$ equation (3.12) has a stationary solution, and if so, what its distributional properties are. This is the main question of this section. Now we give a precise definition of a stochastic recurrence equation in the multivariate setting.

**Definition 3.2.1** We say that a $d$-dimensional stochastic process $(X_t)$ satisfies a stochastic recurrence equation (SRE) or is the solution to a SRE, if there is an i.i.d. sequence $((A_t, B_t))$, where the $A_t$’s are random $d \times d$-matrices and the $B_t$’s are random vectors such that

$$X_t = A_t X_{t-1} + B_t, \quad t \in \mathbb{Z}.$$  \hfill (3.13)
Clearly, (3.13) is nothing but the defining equation of an autoregressive process with random coefficient matrix $A_i$. In the particular case of deterministic $A_i$’s, $(X_t)$ is indeed an autoregressive process and therefore processes satisfying (3.13) are sometimes called generalized (multivariate) autoregressive processes. Various non-linear models can be studied in the context of SREs. Let us consider a few of them.

**Example 3.2.2**
(a) AR(1) process. Assume that, for an i.i.d. sequence $(Z_n)$, the process $(X_n)$ satisfies the equation

$$X_n = \alpha X_{n-1} + Z_n,$$

where $\alpha$ is a fixed constant. Obviously this relation has the same form as (3.13), here $\alpha$ corresponds to the matrix $A_i$ in (3.13).

(b) GARCH(1,1) process. Suppose that random variables $Z_n$, $n \in \mathbb{N}$ are i.i.d.

![Figure 3.2.3](image_url) _A simulated AR(1) time series with $\alpha = 0.6$. The noise $(Z_n)$ is i.i.d. standard normal._

and satisfy the following conditions:

$$E Z_1 = 0 \quad \text{and} \quad \text{Var } Z_1^2 = 1.$$  \hspace{1cm} (3.15)

These conditions are for convenience only, they are chosen in order to make the model computationally less involved. We say that $(X_n)$ is a GARCH(1,1) process,
if
\[ X_n = \sigma_n Z_n, \quad n \in \mathbb{N}, \]

where the \( \sigma_n^2 \)'s satisfy the recursive relation
\[ \sigma_n^2 = \alpha_0 + \alpha_1 X_{n-1}^2 + \beta_1 \sigma_{n-1}^2, \quad n \in \mathbb{Z}. \]

The recursive relation above can be rewritten as a SRE of the form (3.12) with
\[ A_n = \alpha_1 Z_{n-1}^2 + \beta_1 \]
and \( B_n = \alpha_0 \).

(c) GARCH\((p,q)\) process. Assume \( (X_n) \) is a solution to the GARCH equations
\begin{align*}
\sigma_n^2 &= \alpha_0 + \alpha_1 X_{n-1}^2 + \beta_1 \sigma_{n-1}^2, \\
A_n &= \alpha_1 Z_{n-1}^2 + \beta_1, \\
B_n &= \alpha_0.
\end{align*}

Figure 3.2.4 A simulated GARCH\((1,1)\) time series with \( \alpha_0 = 1, \alpha_1 = 0.8 \) and \( \beta_1 = 0.25 \). The noise \( (Z_n) \) is i.i.d. standard normal.

(3.8) and (3.10). For the i.i.d. innovations \( Z_n \) we assume that their common mean and variance are 0 and 1, respectively. One possibility for \( (X_n) \) to be embedded in a SRE is the following: define
\[ X_t = (X_t^2, \ldots, X_{t-p+1}^2, \sigma_t^2, \ldots, \sigma_{t-q+1}^2)' , \]
and notice that this \((p + q)\)-dimensional process satisfies the SRE (3.13) with

\[
A_t = \begin{pmatrix}
\alpha_1 Z_{i}^2 & \cdots & \alpha_{p-1} Z_{i}^2 & \alpha_p Z_{i}^2 & \beta_1 Z_{i}^2 & \cdots & \beta_{q-1} Z_{i}^2 & \beta_q Z_{i}^2 \\
1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\
\alpha_1 & \cdots & \alpha_{p-1} & \alpha_p & \beta_1 & \cdots & \beta_{q-1} & \beta_q \\
0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0
\end{pmatrix},
\]

\[
B_t = (\alpha_0 Z_{i}^2, 0, \ldots, 0, \alpha_0, 0, \ldots, 0)^T. \tag{3.16}
\]

(d) Generalized AR(p) process. Suppose \((X_t)\) is a solution to the equation (3.11). Then we define

\[
X_t = (X_t, \ldots, X_{t-p+1})^T
\]

which again satisfies the SRE (3.13), with the matrix

\[
A_t = \begin{pmatrix}
\Phi_{1}^{(i)} & \Phi_{2}^{(i)} & \cdots & \Phi_{i-2}^{(i)} & \Phi_{i-1}^{(i)} & \Phi_{i+1}^{(i)} \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0
\end{pmatrix}. \tag{3.17}
\]

and the vector

\[
B_t = (Z_t, 0, \ldots, 0)^T. \tag{3.18}
\]

(e) Bilinear process. Assume \((X_n)\) satisfies (3.7). It has a Markovian representation in terms of a SRE too. If \((X_t)\) is a bilinear process then there is a sequence of vectors \((Y_t)\) which satisfies the SRE (3.13) with \(X_t = Y_t\), for some i.i.d. sequence \((A_t, B_t)\) such that \(X_t = Y_t^{(1)} + Z_t\), see Subba Rao and Gabr [64]. Here, \(Y_t^{(1)}\) denotes the first coordinate of the vector \(Y_t\). We only treat the case of the simple bilinear process here (the extremal behavior of such processes is described in Turkman and Turkman [66]). Define \(Y_t = (a + bZ_t)X_t\), and observe that the \(Y_t\)'s satisfy the SRE

\[
Y_t = A_t Y_{t-1} + B_t.
\]
with $A_t = a + bZ_t$ and $B_t = A_t Z_t$. Observe that the pairs $(A_t, B_t)$ form an i.i.d. sequence and $X_t = Y_{t-1} + Z_t$.

Of course it is not a priori clear for which distributions of $(A, B)$ a strictly stationary solution to (3.13) exists. There are various results concerning necessary or sufficient conditions for stationarity, see for example Kesten [35], Vervaat [68] and Bougerol and Picard [9].

Below we give a sufficient condition which remains valid for ergodic sequences $((A_n, B_n))$ (see Brandt [10]) and which is close to necessity (see Babillot et al. [3]).

Before we can formulate this result we need the notion of Lyapunov exponent. By $| \cdot |$ we denote any norm in $\mathbb{R}^d$, and by $\| \cdot \|$ the corresponding operator norm, i.e. for any $d \times d$-matrix $A$,

$$\| A \| = \sup_{|x| = 1} |Ax|.$$

For an i.i.d. sequence $(A_n)$ of i.i.d. $d \times d$ matrices, the constant

$$\gamma = \inf \left\{ \frac{1}{n} \mathbb{E} \ln \| A_1 \cdots A_n \|, \quad n \in \mathbb{N} \right\}$$

(3.19)

is called the top Lyapunov exponent associated with $(A_n)$. Suppose that $\mathbb{E} \ln^+ \| A_1 \| < \infty$. An application of the subadditive ergodic theorem (see Kingman [36]) or results in Funstenberg and Kesten [28] yield

$$\gamma = \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \ln \| A_1 \cdots A_n \| \quad \text{a.s.}$$

(3.20)

Unfortunately, in many interesting cases $\gamma$ cannot be calculated explicitly when $d > 1$. Then we may use relation (3.20) to estimate the value of $\gamma$ via Monte-Carlo simulations of the random matrices $A_n$. By the work of Goldsheid [30] one can even give asymptotic confidence bands for these estimators through a central limit theorem.

Now we are ready to give sufficient conditions for the existence of a stationary solution to a stochastic recurrence equation. The proof is standard. We give it for completeness.

**Theorem 3.2.5** Assume $\mathbb{E} \ln^+ |B_t| < \infty$ and $\gamma < 0$. Then the series

$$X_n = \sum_{m=0}^{\infty} A_n \cdots A_{n-m+1} B_{n-m}$$

(3.21)

converges a.s., and the so-defined process $(X_n)$ is the unique strictly stationary solution of (3.13).
3.2. Stochastic recurrence equations

**Proof.** After a few backward iterations of the SRE (3.13) it becomes clear that the infinite series in (3.21) is a potential candidate for a stationary solution. Its distribution is preserved after an application of SRE (3.13). Therefore it remains to show that the random series in (3.21) converges a.s.

We consider the norm of the general term of the series on the right-hand side of (3.21)

\[ |A_{n} \cdots A_{n-m+1} B_{n-m}| = \exp(m \frac{1}{m} \ln |A_{n} \cdots A_{n-m+1} B_{n-m}|) \]

\[ \leq \exp(m \left( \frac{1}{m} \ln \| A_{n} \cdots A_{n-m+1} \| + \frac{1}{m} \ln |B_{n-m}| \right)) . \]

Use now \( E \ln^+ |B_t| < \infty \) and the strong law of large numbers to see that we have \( \ln^+ |B_{n-m}|/m \to 0 \), as \( m \to \infty \). Moreover,

\[ \frac{1}{m} \ln \| A_{n} \cdots A_{n-m+1} \| \leq \frac{1}{m} \sum_{i=0}^{m-1} \ln \| A_{n-i} \| \overset{a.s.}{\to} \gamma , \]

by the strong law of large numbers. Therefore, with probability 1, there is an \( m_1 \in \mathbb{N} \) such that for every \( m > m_1 \) the right-hand side of the inequality above is less than \( e^{m \gamma /2} \). Hence the series in (3.21) absolutely converges with probability 1.

Strict stationarity of this solution is straightforward to show. Uniqueness follows after iterating forward any other solution, \((Y_n)\) say. After \( m \) steps we have

\[ X_m - Y_m = A_m \cdots A_1 (X_0 - Y_0) \overset{P}{\to} 0 , \]

using the same reasoning as above. Hence \( Y_m \overset{d}{\to} X_0 \) as \( m \to \infty \). Since \((Y_m)\) is strictly stationary its marginal distribution has to be equal to that of \((X_m)\).

\[ \square \]

The solution in (3.21) is causal (or nonanticipative) in the sense that \( X_n \) is independent of the future values of the random parameters \((A_i, B_i)_{i>n}\). It is interesting to mention that if a non-anticipative solution to (3.13) exists and some irreducibility condition holds, then we have \( \gamma < 0 \). We refer to Bougerol and Picard [9] for the proof of this reverse implication.

Notice that by definition the Lyapunov exponent associated with \((A_n)\) is less than zero, as soon as \( E \ln \| A_i \| < 0 \) (take \( n = 1 \) in the definition of \( \gamma \)). If this is true together with \( E \ln^+ |B_t| < \infty \) a stationary solution to (3.13) exists. For the dimension \( d = 1 \) the conditions of Theorem 3.2.5 are much simpler since

\[ \gamma = \frac{1}{n} E \ln |A_1 \cdots A_n| = E \ln |A_1| . \]

We summarize the case \( d = 1 \) as follows:
Corollary 3.2.6 Assume \( d = 1 \), \(-\infty \leq E \ln |A_t| < 0 \) and \( E \ln^+ |B_t| < \infty \). Then the unique stationary solution of (3.13) is given by (3.21).

3.2.2 The strong mixing condition

In Section 2.2 we presented quite general conditions for a Markov chain \((X_n)\) to be strongly mixing. Our intention is to apply these results to the solutions to a SRE. Recall that the Markov chain \((X_n)\) is geometrically ergodic if there is a \( \rho \in (0, 1) \) such that its \( n \)-step transition probabilities with an arbitrary initial point \( y \) satisfy

\[
\rho^{-n} \| p^n(y, \cdot) - \pi(\cdot) \|_{TV} \to 0,
\]

where \( \| \cdot \|_{TV} \) denotes the total variation norm and \( \pi \) is the invariant measure of the Markov chain. For a given random matrix \( A \) we define the function

\[
h_A(v) = E \| A \|^n,
\]

for all the values \( v \in [0, \infty) \). If \( E \| A \|^n < \infty \) for some \( u > 0 \), \( h_A \) is a well defined real function on the interval \([0, u]\). Observe that it has first and second derivatives

\[
h'_A(v) = E \| A \|^n \ln \| A \|
\]

and

\[
h''_A(v) = E \| A \|^n \ln^2 \| A \|.
\]

Since \( h''_A > 0 \), \( h_A \) is a convex function on \([0, u]\).

Lemma 3.2.7 The condition

\[
h_A(\varepsilon) < 1
\]

is satisfied for some value \( \varepsilon > 0 \) if and only if

\[
E \ln \| A \| < 0 \quad \text{and} \quad h_A(\delta) < \infty
\]

for some \( \delta > 0 \).

Proof. The proof is intuitively clear from Figure 3.28.

Indeed, assume \( E \ln \| A \| < 0 \) and \( E \| A \|^\delta < \infty \). Then \( h_A \) is a real function on the interval \([0, \delta]\) which has derivative (from the right) \( h'_A(0+) < 0 \). We immediately observe that \( h(v) \) decreases in a small neighbourhood of zero, and since \( h_A(0) = 1 \) it follows that \( h_A(\varepsilon) < 1 \) for some small \( \varepsilon > 0 \).

On the other hand, \( h_A(\varepsilon) < 1 \) for some \( \varepsilon > 0 \) implies that \( E \ln \| A \| < 0 \) by an application of Jensen’s inequality. \( \Box \)
3.2 Stochastic recurrence equations

Figure 3.2.8 A possible graph of $h_A$ when the derivative $h'_A(0^+)$ is strictly negative.

Proposition 3.2.9 Assume that the distribution of either $A$ or $B$ has a density. Suppose there exists an $\varepsilon > 0$ such that

$$h_A(\varepsilon) < 1 \quad \text{and} \quad E|B|^\varepsilon < \infty. \quad (3.22)$$

Then there exists a unique stationary solution to (3.13) and the Markov chain $(X_n)$ is geometrically ergodic and therefore strongly mixing with geometric rate.

Proof. The existence of a stationary solution follows at once from Theorem 3.2.5 and Lemma 3.2.7. Alternatively, one can establish that the infinite series in (3.21) converges a.s. by showing that the sum of the $\varepsilon$th moments of the summands is finite.

To prove geometric ergodicity of $(X_n)$ we apply Theorem 2.2.2. An application of Lebesgue dominated convergence ensures that for any bounded continuous function $g$, $E(g(X_1)|X_0 = x)$ is continuous in $x$ and hence the Markov chain is Feller, see the comment on page 45.

The $\mu$-irreducibility for some measure $\mu$, follows from the fact that either $A$ or $B$ have a density.

According to Lemma 3.2.7, without loss of generality we may assume $\varepsilon$ in (3.22) lies in the interval $(0,1]$. Now set

$$g(x) = |x|^\varepsilon + 1, \quad x \in \mathbb{R}^d,$$
for \( \varepsilon \in (0, 1] \) given in the assumptions. If there exists a compact \( C \) such that
\[
E(g(X_n) \mid X_{n-1} = x) \leq (1 - \delta)g(x), \quad x \in C^c, \tag{3.23}
\]
the geometric ergodicity and strong mixing property of \((X_n)\) follow from Theorem 2.2.2. Obviously,
\[
E(g(X_n) \mid X_{n-1} = x) \leq E|Ax|^\gamma + E|B|\gamma + 1
\leq E\|A\|\gamma |x|^\gamma + E|B|\gamma + 1
=: E\|A\|\gamma g(x) + L.
\]
Choose \( C \) as the closed ball in \( \mathbb{R}^d \) with center 0 and radius \( M > 0 \) so large that
\[
E(g(X_n) \mid X_{n-1} = x) \leq (1 - \delta)g(x), \quad |x| > M,
\]
for some constant \( 1 - \delta > E\|A\|\gamma \). This proves the desired relation (3.23) and concludes the proof. \( \square \)

We want to see now what the conditions of Theorem 3.2.5 and Proposition 3.2.9 actually mean for some of the non-linear processes of Example 3.2.2.

Example 3.2.10
(a) AR(1). It is straightforward to see that \( \gamma = \log |\alpha| \) for this process. Moreover, condition \( \gamma < 0 \) is equivalent to the well-known condition for the existence of the causal stationary solution to an AR(1) equation namely \( |\alpha| \) has to be less than 1. The stationary solution is geometrically ergodic, as follows by an application of Proposition 3.2.9.
(b) GARCH(1,1). This model is embedded in a 1-dimensional SRE, with \( A_n = \alpha_1 Z_{n-1}^2 + \beta_1 \) as in Example 3.2.2. By Jensen’s inequality, \( E A < 1 \) implies that \( \gamma < 0 \). So in the case \( \text{Var} Z = 1 \), we have that
\[
\alpha_1 + \beta_1 < 1
\]
implies the existence of a stationary solution (as proved by Nelson [46]). The set of pairs \((\beta_1, \alpha_1)\) for which the stationary solution exists is somewhat larger.

Assume that the common distribution of the \( Z_n \)'s is the standard normal. We applied an approximative numerical method to plot this set in Figure 3.2.10. This is the set of all \((\beta_1, \alpha_1)\) in \((0, \infty)^2\) satisfying
\[
E\ln(\alpha_1 Z_{n-1}^2 + \beta_1) < 0 \tag{3.24}
\]
It is possible to describe this set in terms of generalized hypergeometric functions and the digamma function (using the Maple software package for instance). The geometric ergodicity and the strong mixing property follow from Lemma 3.2.7.
Figure 3.2.11  The set of pairs \((\beta, \alpha)\) for which the GARCH(1,1) equations have a stationary solution is bounded by the axes and the upper curve which represents the points for which the expectation in (3.24) is exactly equal to 0. The lower curve correspond to the set \(\beta_1 + \alpha_1 = 1\) which is known (by work of Bougerol and Picard [9]) to be included in this set too.

(c) Simple bilinear process. We have \(\gamma < 0\) whenever \(E \ln |a + bZ| < 0\). We again assume that the sequence \((Z_n)\) has a common standard normal distribution. As for GARCH(1,1), it is possible to give the set of constants \((a, b)\) for which the stationary solution exists. The strong mixing property follows from Lemma 3.2.7, in this case as well.

(d) GARCH\((p, q)\) and generalized AR\((p)\) processes. It is in general very difficult to check in the multidimensional case if the conditions of Theorem 3.2.5 are satisfied. However, in the case when the matrix \(A\) is deterministic, we can calculate \(\gamma\) explicitly:

\[
\gamma = \ln \varphi(A),
\]

(3.25)

where \(\varphi(A)\) denotes the spectral radius of the matrix \(A\). This follows from equation (3.20) and the well known spectral radius formula, see for instance Dunford and Schwartz [24],

\[
\varphi(A) = \lim_{n \to \infty} ||A^n||^{1/n}.
\]
Notice that after taking the logarithm of the right-hand side this is equal to the expression in (3.20). Hence a stationary solution exists provided that the spectral radius of $A$ is less than 1.

(e) Now consider a standard AR($p$) process which satisfies

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t$$

(3.26)

and note that it can be represented as a generalized AR($p$) process with constant coefficient matrix $A_\alpha$. Here the condition $\gamma < 0$ is equivalent to the well known condition for the existence of a stationary causal (i.e. non-anticipative) AR($p$) process. Indeed, assume that ($Z_t$) in (3.26) is a white noise sequence. We know that a stationary causal solution to the equation (3.26) exists if and only if all the roots of the polynomial

$$1 - \alpha_1 z - \cdots - \alpha_p z^p$$

lie outside the unit circle (see Brockwell and Davis [12], Section 3.1), which is another way of asking that the spectral radius of $A_\alpha$ is smaller than 1, or equivalently that $\gamma < 0$.

3.2.3 Regularly varying solutions

Until now, we have established sufficient conditions for the existence of the stationary solution to a given SRE. We also found which conditions have to be satisfied for the strong mixing property. For the asymptotic analysis it would be desirable if we could say something more specific about its distributional properties. It turns out that, under quite general conditions, the stationary solution is regularly varying. These results were successfully applied to study the extremes of solutions to SRE already in de Haan et al. [22].

From now on, we assume that the matrices $A_t$ and the vectors $B_t$ have all entries non-negative. The results we present can be extended to the case of possibly negative entries, however the theory is more involved and additional conditions are needed, we refer to Kesten [35] and LePage [38].

We state here a modification of Kesten’s fundamental result (combined from Theorems 3 and 4 in Kesten [35]). By $g(C)$ we denote the spectral radius of the matrix $C$, while $C > 0$ means that all entries of this matrix are positive. For each matrix $A = (A_{ij})$ in $\mathbb{R}^{d \times d}$, we first define the following value

$$Q_A = \min_{j=1,\ldots,d} \sum_{i=1}^d A_{ij}/\sqrt{d}.$$  

(3.27)

**Theorem 3.2.12** Let $((A_n, B_n))$ be an i.i.d. sequence of the coefficients of a SRE with all entries non-negative. Assume that the i.i.d. sequence of $d \times d$ matrices $A_n$ has joint distribution $P_A$ satisfying:
3.2. Stochastic recurrence equations

(i) For some \( \varepsilon > 0 \), \( E\|A\|^\varepsilon < 1. \)

(ii) \( A \) has no zero rows a.s.

(iii) The set

\[
\{ \ln g(a_n \cdots a_1) : n \in \mathbb{N}, \ a_n \cdots a_1 > 0 \text{ and } a_1, \ldots, a_n \in \text{support of } P_A \}
\]

generates a dense group in \( \mathbb{R} \).

(iv) There exists a \( \kappa_0 > 0 \) such that

\[
E Q_A^{\kappa_0} \geq 1
\]

and

\[
E\|A\|^{\kappa_0} \ln^+ \|A\| < \infty.
\]

Then there exists a unique solution \( \alpha \in (0, \kappa_0] \) to the equation

\[
0 = \lim_{n \to \infty} \frac{1}{n} \ln E\|A_n \cdots A_1\|^{\alpha}.
\]

Furthermore, assume that the i.i.d. sequence \( (B_n) \) satisfies

(v) for the same \( \alpha \)

\[
E|B|^\alpha < \infty.
\]

Then there exists a unique stationary solution \( (X_n) \) to the SRE in (3.13). Moreover, for each vector \( x \in \mathbb{R}^d \), \( x \neq 0 \), the following limit

\[
\lim_{t \to \infty} t^\alpha P((X_1, x) > t)
\]

exists and is finite and strictly positive on \( \mathbb{S}^{d-1} \). This implies (see remark below) in the case of \( \alpha \) which is not an even integer that \( X_1 \) is a regularly varying random vector and moreover the sequence \( (X_n) \) is regularly varying.

Remark 3.2.13 Notice that the relation (3.32) is actually stronger than the regular variation condition for non-negative random vectors. Indeed, condition (3.32) implies that for all \( x \neq 0 \) the following limit exists

\[
\lim_{t \to \infty} \frac{P((X_1, x) > t)}{P((X_1, 1) > t)}.
\]

We may now apply Theorem 2.1.14 to show that the regular variation condition holds in the case when \( \alpha > 0 \) is not an even integer.
The regular variation property for the sequence \((X_n)\) follows once we observe that we can write

\[
(X_m, \ldots, X_0) = (A_m \cdot A_{m-1} \cdots A_1, A_{m-1} \cdots A_1, \ldots, A_1, I)X_0 + B_{(m)}
\]

\[
= A_{(m)}X_0 + B_{(m)}.
\]

where \(B_{(m)}\) denotes the remainder term. It follows from Proposition 2.1.18 that the vector \(A_{(m)}X_0\) is regularly varying while \(B_{(m)}\) does not contribute to the asymptotic behaviour of the tail of \((X_m, \ldots, X_0)\). This follows directly from Remark 2.1.20.

**Remark 3.2.14** A direct consequence of Kesten's Theorem 3.2.12 is that the stationary solution of the SRE satisfies the \(A(a_n)\) condition. Indeed, the strong mixing condition follows by Proposition 3.2.9, which, by an application of Lemma 2.3.9, further implies \(A(a_n)\). Therefore we can apply the results of Section 2.3.

**Remark 3.2.15** The condition (iv) of Theorem 3.2.12, and the inequality (3.29) in particular, are difficult to check directly. However, (3.29) is implied by the simpler condition:

\[
P(Q_A > 1) > 0.
\]

In the case \(d = 1\), Goldie [29] gave an alternative proof of the regular variation property for solutions to a SRE. The conditions for the regular variation of \((X_n)\) are also simpler in this case, which is the reason we state this case as a separate theorem.

**Theorem 3.2.16** Assume that the sequence \(((A_t, B_t))\) of 2-dimensional i.i.d. random vectors satisfies for some \(\alpha > 0\) the following conditions

(a)

\[
E|A|^\alpha = 1, \quad E|A|^\alpha \log^+ |A| < \infty,
\]

(b) the conditional law of \(\log^+ |A|\), given \(A \neq 0\), is nonarithmetic and

(c)

\[
E|B|^\alpha < \infty.
\]

Then the stationary solution to

\[
X_t = A_t X_{t-1} + B_t
\]
exists and the tails of the distribution function of \( X \) are approximately power laws, namely

\[
P(X > t) \sim C_+ t^{-\alpha} \quad \text{as } t \to \infty,
\]

\[
P(X < t) \sim C_- t^{-\alpha} \quad \text{as } t \to \infty,
\]

for some non-negative constants \( C_+ \), \( C_- \) which are both equal to zero if and only if for some \( c \in \mathbb{R} \) we have \( P(B = c(1 - A)) = 1 \). Otherwise, the process \( (X_t) \) is regularly varying.

Goldie [29] gives in addition to this theorem expressions for the constants \( C_- \) and \( C_+ \). From his Lemma 2.2 and our Lemma 3.2.7 we also conclude that a random variable \( A \) which satisfies the conditions of Theorem 3.2.16 also satisfies assumption (i) of Theorem 3.2.12. This allows us to apply Remark 3.2.14 and to obtain the strong mixing property for the process \( (X_t) \).

![Figure 3.2.17](image)

**Figure 3.2.17** Dependence of the regular variation index \( \alpha \) on the choice of the parameters \( (\beta_1, \alpha_1) \), for the stationary squared GARCH(1,1) process and in the case of standard normal random variables \( Z_n \). The solid line (see also Figure 3.2.8) is the boundary of the set of parameters for which the GARCH(1,1) equation has a stationary solution. Note that the \( X_t \)'s are regularly varying as well, only with the index \( 2\alpha \).
Let us consider some implications of Kesten’s [35] and Goldie’s [29] results for some particular processes in Example 3.2.2. We consider the 1-dimensional processes first.

Following Example 3.2.2, we can write the GARCH(1,1) equation in terms of a SRE with random parameters $(A_n, B_n)$ defined as

$$A_n = \alpha_1 Z^2_{n-1} + \beta_1, \quad \text{and} \quad B_n = \alpha_0.$$ 

Assume there is an $\alpha$ such that

$$E|\alpha_1 Z^2 + \beta_1|^\alpha = 1.$$  \hspace{1cm} (3.36)

Hence the conditions of Theorem 3.2.16 are satisfied if the random variable $Z$ has the standard normal distribution for instance. The index $\alpha > 0$ of regular variation for the random variables $X^2$ is determined by the equation (3.36). The way $\alpha$ depends on the coefficients $(\beta_1, \alpha_1)$ is illustrated in Figure 3.2.17.

Very similar results can be obtained in the case of a simple bilinear process (see Basrak et al. [4]) if we assume that the i.i.d. random variables $Z_n$ have sufficiently many moments and that there exists an $\alpha$ such that

$$E|a + bZ|^\alpha = 1.$$ 

The random coefficients have the following form:

$$A_n = a + bZ_n, \quad \text{and} \quad B_n = a Z_n + bZ^2_n.$$ 

Therefore a stationary simple bilinear process is regularly varying if the i.i.d. sequence $(A_n, B_n)$ satisfies conditions (3.34) and (3.35). The index of regular variation is equal to $\alpha$.

In the case of multivariate processes we are in general not able to give a simple equation for the $\alpha$ of Theorem 3.2.12. However, it is possible to check the sufficient conditions of that theorem and to establish the existence of such an $\alpha$.

Suppose that the process $(X_t)$ is a non-anticipative stationary GARCH($p,q$) process. Therefore, Lyapunov’s exponent $\gamma$ of the matrices $(A_t)$ is strictly negative. Recall that the random matrices $A_t$ and the random vectors $B_t$ have the following
form

$$A_t = \begin{pmatrix}
\alpha_1 Z_t^2 & \cdots & \alpha_{p-1} Z_t^2 & \alpha_p Z_t^2 & \beta_1 Z_t^2 & \cdots & \beta_{q-1} Z_t^2 & \beta_q Z_t^2 \\
1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\
\alpha_1 & \cdots & \alpha_{p-1} & \alpha_p & \beta_1 & \cdots & \beta_{q-1} & \beta_q \\
0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0
\end{pmatrix},$$

$$B_t = (\alpha_0 Z_t^2, 0, \ldots, 0, \alpha_0, 0, \ldots, 0)^\top.$$

The following proposition for a stationary GARCH process follows from Theorem 3.2.12.

**Proposition 3.2.18** Assume that the matrix **A** above satisfies

$$E\|A\|^\epsilon < 1$$

for some $\epsilon > 0$

and that $Z$ has a positive density on $\mathbb{R}$ such that

$$E|Z|^h < \infty$$

for all $h \in I$,

where $I = [0, \infty)$ or $I = [0, h_0)$ for some $h_0 > 0$, such that $E|Z|^{h_0} = \infty$.

Then there exists a stationary solution $(X_t)$ to the GARCH($p, q$) equations. Moreover there is an $\alpha > 0$ such that the sequence $(X_t)$ is regularly varying with index $\alpha$.

**Proof.** The existence of a stationary solution (together with the strong mixing property) follows directly from Proposition 3.2.9. To prove the regular variation property, we will use the fact that this property is preserved under power transformations, see Lemma 2.1.21. In particular, if $(Y_1, \ldots, Y_d)$ is regularly varying with index $\kappa$ in $\mathbb{R}_+^d$, then the same holds for the vector $(\sqrt{Y_1}, \ldots, \sqrt{Y_d})$ with index $2\kappa$.

We want to apply Kesten’s Theorem 3.2.12. Since the conditions (i) and (ii) of that theorem are satisfied, we have to check the remaining (iii),(iv) and (v).

We first iterate the original SRE $m$ times (see (c) of Example 3.2.2). Notice that the sequence $(X_t)$ has a subsequence, namely $(X_{t+m})$, which satisfies the following SRE

$$X_{t+m} = A_{t+m} \cdot A_{t+m-1} \cdots A_t X_t + B_{(t+m)}$$

$$= \bar{A}_{t+m} X_0 + \bar{B}_m.$$
For $m$ big enough each entry of the matrix $\tilde{A}_{t+m}$ is positive with probability 1. This follows from the fact that each entry $\tilde{A}_{t+m}^{(t+m)}$ of the matrix $\tilde{A}_{t+m}$ is a sum of terms of the following form

$$
\sum_{m_1, \ldots, m_k} c_{i,j}^{m_1, \ldots, m_k} Z_{m_1}^2 \cdots Z_{m_k}^2 + c_{i,j}^t
$$

where the indices of summation $m_1, \ldots, m_k$ satisfy $t \leq m_1 < \cdots < m_k \leq t + m$, $k \leq m$, while $c_{i,j}^t, k \leq m$ are constants not all equal to 0. Moreover, each entry is a continuous function of $(Z_t, \ldots, Z_{t+m})$. Therefore the support of $\tilde{A}_{t+m}$ is a connected set.

Since the function $\ln g(A)$ is a continuous function of the entries of the matrix $A$, this means that the set in (3.28) is connected. It clearly contains at least two points, although the proof of this fact might be tiresome. Hence it contains an interval connecting these two points and therefore the group generated by this set is dense in $\mathbb{R}$, i.e. the condition (iii) of Theorem 3.2.12 is satisfied.

Finally, notice that, by the choice of the index $m$, we have for $Q_A$ given by (3.27)

$$
P(Q_A > 1) > 0.
$$

Hence Remark 3.2.15 gives the part (3.29) of the condition (iv). On the other hand, the moments in condition (3.30) and in condition (v) are finite because of the assumption on the moments of $Z$.

Therefore, all the conditions of Theorem 3.2.12 are established and the statement of the proposition follows for the sequence $(X_t)$. Recall that it consists of the squares of the original GARCH($p, q$) process as

$$
X_t = (X_t^2, \ldots, X_{t-p+1}^2, \sigma_t^2, \ldots, \sigma_{t-q+1}^2)'.
$$

But the distribution of $(X_t, \ldots, X_{t-m})$ is equal to that of the following random vector

$$
(\epsilon_t \sqrt{X_t^2, \ldots, \epsilon_{t-m} \sqrt{X_{t-m}^2}},
$$

for i.i.d. Rademacher random variables $\epsilon_t$. Recall that $\epsilon$ takes values $\pm 1$ with probability 1/2.

Now apply Lemma 2.1.21 to the random vector $(X_t^2, \ldots, X_{t-m}^2)$, which is regularly varying, with index $\kappa$ say. Finally, the sequence $(X_t, \ldots, X_{t-m})$ is regularly varying with index $\alpha = 2\kappa$, since the Rademacher random variables have moments of all orders and we can apply Corollary 2.1.19. \qed