Lecture Notes in Statistics 181

Daniel Straumann

Estimation in Conditionally Heteroscedastic Time Series Models

Springer

Lecture Notes in Statistics **181**
Edited by P. Bickel, P. Diggle, S. Fienberg, U. Gather,
I. Olkin, S. Zeger

Daniel Straumann

Estimation in Conditionally Heteroscedastic Time Series Models

Springer

Daniel Straumann
Ottikerstrasse 40
CH-8006 Zürich
Switzerland
strauman@math.ethz.ch

Library of Congress Control Number: 2004115047

ISBN 3-540-21135-7 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in Germany

Cover design: *design & production*, Heidelberg
Typesetting by the author using LaTeX
Printed on acid-free paper 40/3142YL-5 4 3 2 1 0

To my parents

Preface

This monograph arose from my PhD thesis, which was written under the supervision of Prof. T. Mikosch at the universities of Groningen and Copenhagen. The extensions and generalizations of the PhD material were carried out at RiskLab, Department of Mathematics, ETH Zurich.

Acknowledgements

I am indebted to my PhD supervisor Prof. T. Mikosch who supported me throughout my PhD time. Without his inputs and continuous criticism this monograph would not exist. I am also grateful for the advice of numerous other people who were kind enough to talk with me about aspects related to this work.

In particular I would like to thank all those who received my work in a positive way. The list starts with the PhD committee: Prof. M. Sørensen (Copenhagen), Prof. F. Liese (Rostock) and Prof. A. McNeil (Zurich). During my year at RiskLab I was invited by Prof. M. Deistler (Vienna), Profs. J.-P. Kreiss and M. Neumann (Braunschweig), Prof. R. Dahlhaus and Dr. S. Subba Rao Tata (Heidelberg) and Prof. U. Schmock (Vienna). On this occassion I would like to express my gratitude for all these invitations and for the extraordinary hospitality that I experienced. The discussions I had were a great source of motivation. In a workshop I had the opportunity to meet "Berkes et al." (i.e. Profs. I. Berkes, L. Horváth and P. Kokoszka). Their paper [8] had inspired me to consider likelihood inference in conditionally heteroscedastic time series models. Furthermore I had email contact with Profs. F. Liese (Rostock) and A. Mokkadem (Paris). Thank you very much for answering all my questions ! I also received positive feedback from Profs. G. Balkema, F. Drost and Q. Yao.

I would like to acknowledge the generous support of RiskLab Zurich without which I could not have improved my PhD thesis. My stay at RiskLab was organized by Profs. P. Embrechts and A. McNeil. Special thanks go to Dr. M. Studer. He read my monograph very carefully and made many suggestions for improvements.

To my parents I owe moral support, and for this reason I dedicate this booklet to them.

Zurich,
September 2004 *Daniel Straumann*

Contents

List of Figures

List of Tables

1

Introduction

1.1 A Very Brief History of Financial Time Series

Conditionally heteroscedastic models for time series play an important role in todays financial risk management, which typically tries to make financial decisions based on observed discrete time asset price data P_t. The prices P_t are believed to be nonstationary, and therefore they are often transformed to so–called log–returns

$$X_t = \log P_t - \log P_{t-1}.$$

The log–returns are approximately equal to the relative price changes and appear to be in agreement with the hypothesis of stationarity, at least over periods of time that are not too long (see Chapter 3).

It has frequently been suggested that (X_t) constitutes a sequence of independent and identically distributed (iid) random variables, or in other words, that the log–prices $(\log P_t)$ evolve according to a random walk. The foundation for the random walk hypothesis was laid by the work of Samuelson and is summarized in his comprehensive survey article [122]. Samuelson proposed modeling speculative prices in continuous time imposing a geometric Brownian motion. The discretization of this model leads to a random walk with iid Gaussian increments for the discrete time log–prices. In the early days of modern finance, the normal distribution played an important role. Its mathematical tractability, particularly in the multidimensional case, paved the way to elegant theories, such as the Markowitz portfolio theory [93] or the option pricing theory of Black and Scholes [16], to name but a few.

For some time, the random walk hypothesis with Gaussian increments was not statistically tested, mainly because of the fact that the graphical representation and analysis of large data sets was very complicated without the assistance of computers. This situation changed in the early sixties. Computers became widespread, and people started to analyze financial data. The hypothesis of geometric Brownian motion was soon rejected, e.g. by Mandelbrot [92] and Fama [48]. Their empirical studies were based on several US stock log–return time series and their conclusion was that

- there is serial dependence in the data,
- the volatility changes over time,
- the marginal distribution of the data is heavy–tailed and asymmetric and hence non–Gaussian.

These observations, sometimes called "stylized facts of financial data" and illustrated in Chapter 3, are clear indications that a random walk with Gaussian increments is not a very realistic model for financial log–price data.

It took some time before a satisfactory discrete time model was found by Engle [46]. This important discovery was recognized by the Nobel Prize in Economics of 2003. Engle's model kills two birds with one stone. Not only does it capture the stylized facts described above relatively well, but it is also simple and in particular stationary, so that statistical inference is possible. Engle avoided the less elegant and more ad–hod methods to allow for shifts in the variance, which were in fashion at that time. He called his model autoregressive conditionally heteroscedastic (ARCH) because its conditional variance (squared volatility) is nonconstant over time and shows an autoregressive structure. In ARCH, the observations are equal to iid white noise (Z_t), up to coordinate–wise multiplication with a positive process (σ_t), where for every fixed t the noise variable Z_t and σ_t are assumed independent. The sequence (σ_t) is also called the volatility process. The dynamics of (σ_t^2) are given via a linear regression with past squared observations.

A few years later, Bollerslev [18] enlarged the ARCH class by the introduction of GARCH (generalized autoregressive conditionally heteroscedastic) models. A stochastic process (X_t) is called GARCH(p,q) (generalized autoregressive conditionally heteroscedastic) if it satisfies the equation

$$X_t = \sigma_t Z_t, \tag{1.1}$$

where (Z_t) is a sequence of iid random variables with $\mathbb{E}Z_0 = 0$ and $\mathrm{Var}(Z_0) = 1$, and (σ_t) is a nonnegative process obeying the recursive equation

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \,. \tag{1.2}$$

The parameters α_i and β_j are nonnegative and the case $q = 0$ corresponds to ARCH(p). Again, σ_t and Z_t are independent for every fixed t. The use of GARCH often enhances parsimony, because compared to a pure ARCH process, less parameters are needed for the description of the data. GARCH(p,q) is still the benchmark model, although a variety of alternative conditionally heteroscedastic time series models have been proposed in the meantime (see Shephard [123] or Carrasco and Chen [31] for references).

1.2 Contents of the Monograph

Despite the seemingly simple model equations (1.1)–(1.2), some of the stochastic properties of GARCH(p, q) are not easy to establish. GARCH differs substantially from the (linear) autoregressive moving average (ARMA) models (see Section 3.2 for their formal definition). This starts already with the conditions for stationarity. When GARCH was introduced, it was quickly assumed that $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ is necessary and sufficient for the existence of a unique stationary solution to (1.1)–(1.2). The "proof" given for this statement was the fact that the squares of a GARCH(p, q) process obey a certain ARMA$(\max(p, q), q)$ equation with noise sequence $(X_t^2 - \sigma_t^2)$, for which $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ is necessary and sufficient for causality. Such a conclusion is however incorrect because it already presupposes the existence of a stationary GARCH(p, q) process. At that time it was furthermore not realized that the noise sequence is not iid and might have an infinite second moment. Nelson [105] found the correct criterion for stationarity in GARCH$(1, 1)$: necessary and sufficient for stationarity is the condition $\mathbb{E}[\log(\alpha_1 Z_0^2 + \beta_1)] < 0$, and there are parameters with $\alpha_1 + \beta_1 > 1$ in the domain of stationarity. Later, Bougerol and Picard [21] gave a sufficient and necessary condition for stationarity of GARCH(p, q); however, as we will see in Section 3.3, this condition is in terms of the top Lyapunov exponent associated with a multidimensional stochastic dynamical system, and it cannot be verified other than with Monte Carlo methods.

Another source of difficulty was the limit behavior of the sample autocovariances of X_t, $|X_t|$ or X_t^2. A complete answer was given by Basrak et al. [5] (see Section 8.2 for the special case of GARCH$(1, 1)$). As regards parameter estimation, many questions remained open until recently, when Berkes et al. [8] studied the so–called quasi maximum likelihood estimator (QMLE) in GARCH(p, q), under minimal assumptions.

This book deals with parameter estimation in conditionally heteroscedastic time series models. We extend the results of Berkes et al. [8] insofar that we allow more general and "highly nonlinear" conditionally heteroscedastic time series models such as asymmetric or exponential GARCH (see Section 1.2.1 below). We base our reasoning on so–called stochastic recurrence equations. We believe that this approach is adapted to the problem and that it leads simultaneously to a high degree of generality, simplicity and clarity. A second part is devoted to questions related to the limit behavior of the QMLE if the tails of the innovations are heavy and of the Whittle estimator in GARCH$(1, 1)$.

We have tried to write up the ideas and techniques in a pedagogical way so that future research can hopefully benefit from our considerations. Although the book rather focuses on theoretical questions and mathematical foundations, many illustrations are included. In what follows, we provide some more specific information about the topics of this book.

1.2.1 Parameter Estimation in a General Conditionally Heteroscedastic Time Series Model

A considerable part of this book (the Chapters 5 – 7) is devoted to the analysis of a multiplicative model of form

$$\begin{cases} X_t &= \sigma_t Z_t, \\ \sigma_t^2 &= g_{\boldsymbol{\theta}}(X_{t-1},\ldots,X_{t-p},\sigma_{t-1}^2,\ldots,\sigma_{t-q}^2), \end{cases} \qquad t \in \mathbb{Z}, \tag{1.3}$$

where the volatility process (σ_t) is nonnegative and $\{g_{\boldsymbol{\theta}}\}$ denotes a finite–dimensional parametric family of nonnegative functions on $\mathbb{R}^p \times [0,\infty)^q$ fulfilling certain conditions. (Z_t) is a sequence of iid random variables with $\mathbb{E}Z_0 = 0$ and $\mathrm{Var}(Z_0) = 1$. We assume that Z_t is independent of σ_t for every fixed t. This fairly large class of models creates conditional heteroscedasticity because $\mathrm{Var}(X_t \mid X_{t-1},\ldots,X_{t-p+1},\sigma_{t-1}^2,\ldots,\sigma_{t-q+1}^2) = \sigma_t^2$ is in general a random process. It contains e.g. GARCH(p,q), asymmetric GARCH(p,q) and exponential GARCH (see Section 3.3 for a definition of these models). In Chapter 5 we study the so–called quasi maximum likelihood estimator (QMLE) and Chapter 6 provides an analysis of the maximum likelihood estimator (MLE) applied to the model (1.3). Immediately, there are important questions which might be raised:

- How can one establish stationarity conditions for model (1.3) ?
- How can one approximate the unobservable sequence (σ_t^2) from data ?
- How can one build up a (quasi) likelihood function ?
- How does one derive the limit properties of the (Q)MLE ?

The aim is to give answers as general as possible, but still useful enough to be applicable to concrete examples. For the sake of simplicity we restrict ourselves to the case $p = q = 1$ in the following exposition. As we will see in the subsequent chapters, it is apart from notational complications not too demanding to generalize these ideas to higher order models.

Stationarity

By replacing X_{t-1} in (1.3) by $\sigma_{t-1}Z_{t-1}$, one recognizes that (σ_t^2) is a homogeneous Markov chain with transition mechanism

$$\sigma_t^2 = g(\sigma_{t-1}Z_{t-1},\sigma_{t-1}^2), \qquad t \in \mathbb{Z}, \tag{1.4}$$

where $g = g_{\boldsymbol{\theta}_0}$ with $\boldsymbol{\theta}_0$ the true parameter. Alternatively, one can interpret relationship (1.4) as a random transformation of σ_{t-1}. Traditionally the stationarity (or stochastic stability) of nonlinear time series models has been treated with techniques from the theory of general state space Markov chains, see e.g. Meyn and Tweedie [96]. We believe that applying this methodology would lead to an estimation theory which is too general, such that a variety of new problems would be created. Likelihood inference for discrete time Markov processes was developed by Billingsley [11] and later extended to non–Markovian discrete time processes by Hall and Heyde [62], but a plain application of these fairly general results to model (1.3) is not straightforward because of the fact that (σ_t) is unobservable.

Instead we use contraction techniques of so–called stochastic recurrence equations (SREs). An example of a SRE is provided by the recursive relationship (1.4). Note that one can alternatively write (1.4) as

$$\sigma_{t+1}^2 = \psi_t(\sigma_t^2), \qquad t \in \mathbb{Z}, \tag{1.5}$$

where $\psi_t(s) = g(\sqrt{s}Z_t, s)$ is a random map $[0,\infty) \to [0,\infty)$. This means that σ_{t+1}^2 is a random transform of σ_t^2. For a definition of the notion of a SRE on a general separable complete metric space, which is due to Bougerol [20], and a summary of the relevant tools, we refer to Section 2.6 of this monograph. Concerning SREs we have been very much inspired by the excellent survey article by Diaconis and Freedman [38]. An important technique in this field is the backward iteration idea, which says that one should start the Markov chain induced by (1.5) at a time point $t-m$ in the past with some initial value, ς_0^2 say. The observation at time t of the Markov chain initialized in this way is called the mth backward iterate and equal to

$$\tilde{\sigma}_{t,m}^2 = \psi_{t-1} \circ \psi_{t-2} \circ \cdots \circ \psi_{t-m}(\varsigma_0^2).$$

The further in the past the Markov chain starts, i.e., the larger m, the "closer" to stationarity will the chain be at the present, i.e., at time t. Letac [86] was the first who made this intuition mathematically precise. His principle can be summarized as follows: if

- $(\tilde{\sigma}_{t,m}^2)_{m\in\mathbb{N}}$ converges almost surely (a.s.), with a limit irrespective of ς_0^2,
- the random map ψ_t is continuous a.s.,

then $(\lim_{m\to\infty} \tilde{\sigma}_{t,m}^2)_{t\in\mathbb{Z}}$ is the unique stationary sequence obeying (1.5). In practice one would like to have mathematically tractable criteria which imply the almost sure convergence of $(\tilde{\sigma}_{t,m}^2)_{m\in\mathbb{N}}$. Similarly to the classical Banach fixed point theorem for deterministic sequences generated by functional iteration, the requirement that ψ_t is "contractive on average", i.e.,

$$|\psi_t(s) - \psi_t(\tilde{s})| \le \Lambda(\psi_t)\,|s - \tilde{s}| \tag{1.6}$$

with $\mathbb{E}[\log \Lambda(\psi_0)] < 0$, suffices for the almost sure convergence of $(\tilde{\sigma}_{t,m}^2)_{m\in\mathbb{N}}$, with a limit irrespective of ς_0^2. Here $\Lambda(\psi_t)$ denotes the random Lipschitz coefficient

$$\Lambda(\psi_t) = \sup_{s,\tilde{s}\geq 0, s\neq\tilde{s}} \frac{|\psi_t(s) - \psi_t(\tilde{s})|}{|s - \tilde{s}|}.$$

The condition $\mathbb{E}[\log \Lambda(\psi_0)] < 0$ can even be weakened. Already the contractivity of r–fold iterates for a certain $r \geq 1$ implies the almost sure convergence of $(\tilde{\sigma}^2_{t,m})_{m\in\mathbb{N}}$ with a limit irrespective of ς_0^2. This means that it is enough to require

$$\mathbb{E}[\log \Lambda(\psi_0 \circ \cdots \circ \psi_{-r+1})] < 0. \tag{1.7}$$

Since Λ is submultiplicative, the condition (1.7) is certainly weaker than $\mathbb{E}[\log \Lambda(\psi_0)] < 0$. Often it is not enough to analyze the maps ψ_t when one wants to establish the almost sure convergence of the sequence $(\tilde{\sigma}^2_{t,m})_{m\in\mathbb{N}}$ as it might well be that $\mathbb{E}[\log \Lambda(\psi_0)] \geq 0$ but $\mathbb{E}[\log \Lambda(\psi_0 \circ \cdots \circ \psi_{-r+1})] < 0$ for a certain r. This is e.g. the case for the SRE arising from the state space representation of GARCH(p,q), see Section 3.3.1. The key result about stationary solutions of SREs used over and over again in this monograph, Theorem 2.6.1, relies on a contraction condition of form (1.7).

Reconstruction of (σ_t)

Now, the problem of stationarity being resolved, we assume data $X_0, \ldots, X_n$ from model (1.3) has been observed. The "canonical approach" used by time series analysts and econometricians in order to approximate σ_t^2 from data $X_0, \ldots, X_n$, is to take a starting value $\hat{\sigma}_0^2 = \varsigma_0^2$ say, and then to iteratively generate

$$\hat{\sigma}^2_{t+1} = \phi_t(\hat{\sigma}^2_t) \tag{1.8}$$

with $\phi_t(s) = g(X_t, s)$ for $t = 1, \ldots, n$ (see e.g. Nelson [106]). To make this algorithm work, the error of approximation should decay to zero as $t \to \infty$, i.e.,

$$|\hat{\sigma}_t^2 - \sigma_t^2| \xrightarrow{\mathbb{P}} 0. \tag{1.9}$$

It is shown in Section 5.2 that this requirement can be interpreted as an invertibility property of the model. Note that an alternative interpretation of (1.8) is that $(\hat{\sigma}_t^2)_{t\in\mathbb{N}}$ is a solution of the SRE

$$s_{t+1} = \phi_t(s_t), \qquad t \in \mathbb{N}. \tag{1.10}$$

Another statement of the key result Theorem 2.6.1 is as follows: if the random map $\phi_0 \circ \cdots \circ \phi_{-r+1}$ is "contractive on average" for some $r \geq 1$, i.e., $\mathbb{E}[\log \Lambda(\phi_0 \circ \cdots \circ \phi_{-r+1})] < 0$, then the so–called forward iterates $(\hat{\sigma}_t^2)_{t\in\mathbb{N}}$ converge towards the stationary solution (σ_t^2) of the SRE (1.10) with index set $\mathbb{Z}$, and the error is exponentially decaying, i.e., there is a $\gamma > 1$ such that

$$\gamma^t \, |\hat{\sigma}_t^2 - \sigma_t^2| \xrightarrow{\text{a.s.}} 0, \qquad t \to \infty.$$

Thus the condition $\mathbb{E}[\log \Lambda(\phi_0 \circ \cdots \circ \phi_{-r+1})] < 0$ for an $r \geq 1$ implies (1.9).

Deriving a Likelihood

Suppose that Z_t has density $k_{\boldsymbol{\nu}}$, where $\boldsymbol{\nu}$ is some nuisance parameter. Then, provided we knew (σ_t), the log–likelihood of observations $(X_1, \ldots, X_n)$ would equal

$$\sum_{t=1}^{n} \left(-\log \sigma_t + \log k_{\boldsymbol{\nu}}(X_t/\sigma_t)\right). \tag{1.11}$$

But now we need to approximate σ_t^2 under the parameter hypothesis $\boldsymbol{\theta}$. The "canonical approach" is similar to the fixed parameter case $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Take an initial value $\hat{h}_0(\boldsymbol{\theta}) = \varsigma_0^2$ and set

$$\hat{h}_{t+1}(\boldsymbol{\theta}) = g_{\boldsymbol{\theta}}(X_t, \hat{h}_t(\boldsymbol{\theta})) \tag{1.12}$$

for $t \geq 1$. This generates a sequence of random functions $(\hat{h}_t)_{t\in\mathbb{N}}$, and by replacing σ_t^2 by $\hat{h}_t(\boldsymbol{\theta})$, we obtain the log–likelihood

$$\hat{L}_n(\boldsymbol{\theta}, \boldsymbol{\nu}) = \sum_{t=1}^{n} \left\{ -\log \sqrt{\hat{h}_t(\boldsymbol{\theta})} + \log k_{\boldsymbol{\nu}} \left(\frac{X_t}{\sqrt{\hat{h}_t(\boldsymbol{\theta})}} \right) \right\}.$$

The MLE is the maximizer of $\hat{L}_n$ with respect to $(\boldsymbol{\theta}, \boldsymbol{\nu})$. Often one does not want to specify the class $\{k_{\boldsymbol{\nu}}\}$. Then it is advisable to maximize the log–likelihood with respect to a standard normal density (i.e., $k_{\boldsymbol{\nu}}(x) = (2\pi)^{-1/2}e^{-x^2/2}$) since the resulting estimator, which is called the (Gaussian) quasi maximum likelihood estimator (QMLE), is strongly consistent under relatively weak regularity assumptions, see Section 5.3.

A Stationary Approximation of $(\hat{\mathbf{h}}_t)_{t\in\mathbb{N}}$

Observe that $(\hat{h}_t)_{t\in\mathbb{N}}$ is nonstationary in general. Before we can tackle the limit properties of the (Q)MLE, we need to present the idea of a stationary approximation. When starting to think about the problem of finding an approximation of $(\hat{h}_t)_{t\in\mathbb{N}}$ by a stationary sequence $(h_t)_{t\in\mathbb{N}}$, one soon arrives at the conjecture that such a sequence must obey

$$h_{t+1}(\boldsymbol{\theta}) = g_{\boldsymbol{\theta}}(X_t, h_t(\boldsymbol{\theta})), \qquad t \in \mathbb{N}. \tag{1.13}$$

Assume $\boldsymbol{\theta} \in K \subset \mathbb{R}^d$, where K is compact, and denote by $\mathbb{C}(K)$ the space of continuous functions on K, equipped with the sup–norm $\|\cdot\|_K$. We then introduce random maps Φ_t on $\mathbb{C}(K)$ given by

$$[\Phi_t(s)](\boldsymbol{\theta}) = g_{\boldsymbol{\theta}}(X_t, s(\boldsymbol{\theta})), \qquad s \in \mathbb{C}(K).$$

Comparing this with (1.13), we deduce that $(h_t)_{t\in\mathbb{N}}$ is a stationary solution of the SRE

$$s_{t+1} = \Phi_t(s_t), \qquad t \in \mathbb{N}, \tag{1.14}$$

on $\mathbb{C}(K)$. We study the existence and uniqueness of a stationary solution to the SRE (1.14) with index set $\mathbb{Z}$ again by means of Theorem 2.6.1. By similar arguments as before, the backward iterates to (1.14) converge almost surely if there is $r \geq 1$ such that the random map $\Phi_0 \circ \cdots \circ \Phi_{-r+1}$ is "contractive on average", i.e., $\mathbb{E}[\log \Lambda(\Phi_0 \circ \cdots \circ \Phi_{-r+1})] < 0$. In that case, the sequence defined by

$$h_t = \lim_{m\to\infty} \Phi_{t-1} \circ \cdots \circ \Phi_{t-m}(\varsigma_0^2), \qquad t \in \mathbb{Z},$$

is the unique stationary ergodic solution of the SRE (1.14) (with index set $\mathbb{Z}$); cf. Section 5.2.3. Moreover, the forward iterates $(\hat{h}_t)_{t\in\mathbb{N}}$ approach $(h_t)_{t\in\mathbb{N}}$ with an error decaying exponentially fast as $t \to \infty$, i.e., there is a $\tilde{\gamma} > 1$ such that

$$\tilde{\gamma}^t \|\hat{h}_t - h_t\|_K \xrightarrow{\text{a.s.}} 0, \qquad t \to \infty. \tag{1.15}$$

Thus the stationary ergodic sequence $(h_t)_{t\in\mathbb{N}}$ indeed approximates $(\hat{h}_t)_{t\in\mathbb{N}}$.

The SRE approach is also useful for the treatment of the first and second derivatives of $\hat{h}_t$. From differentiation of both sides of (1.13) with respect to $\boldsymbol{\theta}$, we obtain

$$\hat{h}'_{t+1}(\boldsymbol{\theta}) = \frac{\partial g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}\,(X_t, \hat{h}_t(\boldsymbol{\theta})) + \frac{\partial g_{\boldsymbol{\theta}}}{\partial s}\,(X_t, \hat{h}_t(\boldsymbol{\theta}))\,\hat{h}'_t(\boldsymbol{\theta}). \tag{1.16}$$

Similar arguments as before yield that $(\hat{h}'_t)_{t\in\mathbb{N}}$ can be approximated by the stationary sequence $(d_t)_{t\in\mathbb{N}}$, where (d_t) is the unique stationary ergodic solution of the linear SRE

$$s_{t+1} = \frac{\partial g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}\,(X_t, h_t) + \frac{\partial g_{\boldsymbol{\theta}}}{\partial s}\,(X_t, h_t)\, s_t, \qquad t \in \mathbb{Z},$$

and $\|\hat{h}'_t - d_t\|_K \to 0$ at an exponential rate when $t \to \infty$. Not unexpectedly it turns out that d_t coincides with the first derivative h'_t. Similar statements hold true for $\hat{h}''_t$, see Section 5.5. It is perhaps worth mentioning that the random maps transforming $\hat{h}'_t$ into $\hat{h}'_{t+1}$ are *non*stationary (since $(\hat{h}_t)_{t\in\mathbb{N}}$ is nonstationary). By virtue of (1.15) these random maps "tend to stationary ones with an error decaying exponentially fast". The additional difficulty arising from nonstationary transformations is treated in Theorem 2.6.4.

The Limit Properties of the (Q)MLE

Our approach to the derivation of the limit properties is classical, see e.g. Lehmann [85], Ferguson [51] or van der Vaart [129]. A brief summary of the main steps in such proofs is given in the following:

Strong Consistency.

One first shows that $\hat{L}_n/n \xrightarrow{\text{a.s.}} L$, where the almost sure convergence is uniform, say on a compact set which contains the true parameter. In a second step, one proves that the limit function L is uniquely maximized at the true parameter $(\boldsymbol{\theta}_0, \boldsymbol{\nu}_0)$; this implies the strong consistency of the (Q)MLE (note that there is no maximization with respect to $\boldsymbol{\nu}$ in the case of a QMLE).

Asymptotic Normality.

The proof relies on a second order Taylor expansion of $\hat{L}_n$ about the (Q)MLE of $(\boldsymbol{\theta}, \boldsymbol{\nu})$. Although the methodology is rather standard, the necessary validation of all the formal arguments is fairly technical.

Difficulties and their Resolution.

The main difficulty stems from the fact that $(\hat{h}_t)_{t\in\mathbb{N}}$ is nonstationary so that the limit of $\hat{L}_n/n$ cannot be obtained via the ergodic theorem. At first sight, the influence of the initializing constant ς_0^2 in the definition (1.12) of $(\hat{h}_t)_{t\in\mathbb{N}}$ is unclear. A way out is offered by the stationary approximation $(h_t)_{t\in\mathbb{N}}$. One replaces $\hat{h}_t$ in (1.11) by h_t and so obtains an approximation

$$L_n(\boldsymbol{\theta}, \boldsymbol{\nu}) = \sum_{t=1}^{n} \left\{ -\log \sqrt{h_t(\boldsymbol{\theta})} + \log k_{\boldsymbol{\nu}} \left(\frac{X_t}{\sqrt{h_t(\boldsymbol{\theta})}} \right) \right\}.$$

of $\hat{L}_n$. Note that L_n is easier to handle than $\hat{L}_n$ because L_n is the partial sum of a stationary ergodic sequence, so that the use of the ergodic theorem is possible. Since the approximation error $\|\hat{h}_t - h_t\|_K$ tends to zero at a geometric rate, one can show that, uniformly in $(\boldsymbol{\theta}, \boldsymbol{\nu})$,

$$\hat{L}_n(\boldsymbol{\theta}, \boldsymbol{\nu}) = L_n(\boldsymbol{\theta}, \boldsymbol{\nu}) + O_{\mathbb{P}}(1). \tag{1.17}$$

From this uniform error bound, which of course is only valid under regularity assumptions on $\{k_{\boldsymbol{\nu}}\}$, one concludes that strong consistency of the maximizer of L_n and strong consistency of the (Q)MLE are equivalent. Since the sequence $(h_t)_{t\in\mathbb{N}}$ is free of the constant ς_0^2, the latter equivalence (1.17) moreover implies that ς_0^2 is asymptotically irrelevant for the (Q)MLE. From a similar error bound on the first derivative of $\hat{L}_n - L_n$, one can prove that it is sufficient to establish asymptotic normality for the maximizer of L_n. The mathematical elaboration of these ideas leads to many technicalities, as is revealed from a glance at Chapters 5 and 6.

There is one mathematical tool which deserves to be mentioned in the introduction: the ergodic theorem for sequences of random functions in $\mathbb{C}(K, \mathbb{R}^{d'})$. Although it allows one to establish uniform convergence of L_n/n in a simple and straightforward way, it seems to be rather neglected in the literature (see Section 2.2).

Misspecification

The distribution of the innovations Z_t does not need to be specified for the QMLE, as the quasi likelihood is built up under the synthetic assumption of standard Gaussian errors. Despite this misspecification, the QMLE is strongly consistent. This is in line with the statistical rule of thumb which says that maximizers of likelihoods with respect to Gaussian errors provide consistent estimates of the parameter of interest. Another example of this kind is multiple regression, where imposing iid Gaussian errors leads to the least–squares estimator, which is consistent even if the errors are in fact non–Gaussian.

The situation is totally different when we base the MLE on the distributional assumption

$$Z_t \text{ iid } \sim k_{\boldsymbol{\nu}}(x)\,dx, \tag{1.18}$$

and $k_{\boldsymbol{\nu}}$ non–Gaussian. Suppose we are wrong, i.e., there is no $\boldsymbol{\nu}$ such that (1.18) applies. Then the $\boldsymbol{\theta}$–parameter of the MLE is inconsistent. Section 6.2 provides a detailed discussion of this problem, and complements and clarifies similar observations made by Newey and Steigerwald [108].

Heavy Tails

As a matter of fact, for asymptotic normality of the QMLE in GARCH(p,q), the innovations must have a finite 4th moment, i.e., $\mathbb{E}Z_0^4 < \infty$. In Chapter 7 the limit behavior of the QMLE is determined in the case $\mathbb{E}Z_0^4 = \infty$. More precisely, one supposes that Z_0^2 is regularly varying with index $\kappa \in (1,2)$, i.e.,

$$\mathbb{P}(Z_0^2 > z) \;=\; L(z)\,z^{-\kappa}, \qquad z \to \infty,$$

where L is a slowly varying function. We defer the formal definition of a slowly varying function to Chapter 7 and content ourselves to mention that $L(z)$ plays the role of a constant (a constant function is slowly varying). Under the above distributional assumption, one can use a central limit theorem for infinite variance stationary ergodic martingale difference sequences of Mikosch and Straumann [103] in order to obtain the limit behavior of the QMLE:

- The QMLE is strongly consistent.
- The rate of convergence is of magnitude $n^{1-1/\kappa}$, and the limit distribution is (multivariate) κ–stable.

This asymptotic result is of qualitative nature since it is of limited use for inference; at the time being, the parameters of the κ–stable limit distribution are not explicitly known. With different techniques, Hall and Yao [63] established the same result.

1.2.2 Whittle Estimation in GARCH(1,1)

Since the squares X_t^2 of GARCH(p,q) obey an ARMA$(\max(p,q),q)$ equation with white noise, one can apply the classical Whittle estimator to the squares of GARCH(p,q). Giraitis and Robinson [56] show in this context that the Whittle estimator for the parameters in the resulting ARMA model is strongly consistent and asymptotically normal provided the marginal distribution of the process has a finite 8th moment. In Chapter 8 we focus on the case when $\mathbb{E}X_0^8 = \infty$ and GARCH(1,1). This case corresponds to various real–life log–return series of financial data. We show that the Whittle estimator applied to squared GARCH(1,1) is consistent as long as the 4th moment is finite and inconsistent when the 4th moment is infinite. Using classical methods from the theory of Whittle estimation, one shows that the rate of convergence and the limit distribution of the Whittle estimator are determined by the convergence rate of the sample autocovariances of the squares X_t^2 and their joint finite–dimensional weak limits. Once all the necessary technicalities have been overcome for proving this, it is straightforward to apply the results of a paper by Mikosch and Stărică [99]. Using the theory of Kesten [72] and Goldie [58] on the tail behavior of solutions of stochastic recurrence equations, Mikosch and Stărică show that for a stationary GARCH(1,1) process (X_t) there exists a so–called tail index $\kappa > 0$ and a positive constant c with

$$\mathbb{P}(|X_0| > x) \sim c\,x^{-\kappa}, \qquad x \to \infty.$$

The case $\mathbb{E}X_0^4 < \infty$ and $\mathbb{E}X_0^8 = \infty$, which we are considering for Whittle estimation, corresponds to a tail index $\kappa \in (4, 8]$. Moreover, Mikosch and Stărică [99]. prove that the sample autocovariances of the squares X_t^2 are strongly consistent estimators of their deterministic counterparts, with a rate of convergence given by $n^{1-4/\kappa}$. Properly normalized, the finite–dimensional vectors of sample autocovariances converge in distribution to some non–Gaussian limit. Altogether we can conclude that the fatter the tail of X_0, i.e., the smaller κ, the slower the rate of convergence of the Whittle estimator to the true parameters. This limit result is again qualitative and not particularly useful for statistical inference because the emerging limit distribution is not easy to handle. Exactly the same results hold for the least–squares estimator because it is asymptotically equivalent to Whittle.

1.3 Structure of the Monograph

To conclude, we give a short description of the individual chapters:

Chapter 2. We discuss certain mathematical tools and concepts: stationarity, ergodicity, uniform strong law of large numbers, matrix norms, weak convergence, exponentially fast almost sure convergence, SREs.

Chapter 3. We give a short survey of linear and (nonlinear) financial time series models and describe their stochastic properties.

Chapter 4. We motivate certain estimators and provide a survey of existing limit results in the literature.

Chapter 5. We analyze the QMLE in the general model (1.3) and apply the resulting theory to specific models. A small simulation study is included.

Chapter 6. We analyze the maximum likelihood estimator (MLE) in model (1.3) and discuss the issue of misspecification. Moreover, the MLE for AGARCH(p, q) with Student t_ν innovations is treated.

Chapter 7. Here we study the QMLE in model (1.3) when the innovations are heavy–tailed, i.e., have an infinite 4th moment. The resulting theory is applied to GARCH(p, q).

Chapter 8. This is a technical chapter, where the limit properties of the Whittle estimator applied to the squares of GARCH$(1, 1)$ are derived.

2

Some Mathematical Tools

This chapter introduces the relevant notation and gives a collection of auxiliary results, which will be needed later in this monograph. It may be skipped upon first reading since in the subsequent chapters we will indicate whenever we make use of one of the results of this chapter. After a discussion of stationarity and ergodicity of sequences of random elements with values in general spaces, we treat the uniform strong law of large numbers. This is followed by a small section on matrix norms. We then give a summary of results related to weak convergence in the space of continuous functions on compact subsets of $\mathbb{R}^d$. Then we introduce a notion of convergence less common in the literature: exponentially fast almost convergence. The last part is about stochastic recurrence equations, which play a crucial role in our exposition on estimation in conditionally heteroscedastic time series models.

2.1 Stationarity and Ergodicity

Stationarity and ergodicity of stochastic processes play an important role in time series analysis. In what follows we recall the definitions of stationarity and ergodicity for sequences of random elements with values in a general space and state two important results which enable one to establish stationarity and ergodicity in a straightforward way. In our notation we closely follow Section 1.4 in the monograph of Krengel [77]. We also refer to Stout [124], where the more elementary case of stationary real sequences is treated.

As a general definition for this monograph,

$$\mathbb{N} = \{0, 1, 2, \ldots\}.$$

For any sequence $(v_t)_{t\in\mathbb{Z}}$ we write (v_t) in abridged form. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ a measurable space. Then a function on Ω taking values in E and being $\mathcal{A}$–$\mathcal{E}$ measurable is called an E–valued random element. A sequence (v_t) of E–valued random elements is called stationary if for all $t \in \mathbb{Z}$, $n \in \mathbb{N}$ and $h \in \mathbb{Z}$,

$$(v_t,\dots,v_{t+n}) \overset{\mathrm{d}}{=} (v_{t+h},\dots,v_{t+h+n}),$$

where the symbol $\overset{\mathrm{d}}{=}$ denotes equality in distribution. A sequence (v_t) of E–valued random elements gives rise to the definition of the canonical space $E^{\mathbb{Z}}$ of doubly infinite sequences in E, i.e.,

$$E^{\mathbb{Z}} = \{\,(e_t)\,|\,e_t \in E \text{ for all } t \in \mathbb{Z}\,\}.$$

By equipping $E^{\mathbb{Z}}$ with the σ–field $\mathcal{E}^{\mathbb{Z}}$ generated by the family of finite–dimensional product cylinders, i.e.,

$$\mathcal{E}^{\mathbb{Z}} = \sigma\Big(\Big\{\prod_{i\in\mathbb{Z}} C_i \,\Big|\, \exists\, n \geq 0 : C_i = E \text{ for } |i| > n \text{ and } C_i \in \mathcal{E} \text{ for } |i| \leq n\Big\}\Big),$$

we obtain the measurable space $(E^{\mathbb{Z}},\mathcal{E}^{\mathbb{Z}})$. Then a sequence (v_t) of E–valued random elements may also be regarded as a random element taking values in $E^{\mathbb{Z}}$ and inducing a measure $\tilde{\mathbb{P}}$ on $(E^{\mathbb{Z}},\mathcal{E}^{\mathbb{Z}})$ by $\tilde{\mathbb{P}}(C) = \mathbb{P}((v_t) \in C)$ for all $C \in \mathcal{E}^{\mathbb{Z}}$.

Ergodicity of sequences of random elements is defined in terms of the canonical space because it involves a measure preserving transformation, which in general has to be constructed on $(E^{\mathbb{Z}},\mathcal{E}^{\mathbb{Z}},\tilde{\mathbb{P}})$. This transformation is the backshift operator τ in $E^{\mathbb{Z}}$, i.e., $\tau((e_t)) = (e_{t+1})$, $t \in \mathbb{Z}$, which is measurable. It can be verified that τ preserves the measure $\tilde{\mathbb{P}}$, i.e., $\tilde{\mathbb{P}}(\tau^{-1}(C)) = \tilde{\mathbb{P}}(C)$ for all $C \in \mathcal{E}^{\mathbb{Z}}$, if and only if (v_t) is stationary. For the following we assume that (v_t) *is* stationary (and hence τ measure preserving). Furthermore we call a set $C \in \mathcal{E}^{\mathbb{Z}}$ invariant if $\tau^{-1}(C) = C$, and we say that (v_t) is ergodic if for all invariant sets C, either $\tilde{\mathbb{P}}(C) = 0$ or $\tilde{\mathbb{P}}(C) = 1$. The definition of stationarity and ergodicity of one–sided sequences $(v_t)_{t\in\mathbb{N}}$ is similar, see Krengel [77].

The standard example of a stationary ergodic sequence of random elements is an iid sequence; note that the ergodicity of an iid sequence is implied by Kolmogorov's 0–1 law. If one has a certain stationary ergodic sequence, it is relatively simple to generate others. This is the content of the following well–known result, see e.g. Proposition 4.3 in Krengel [77].

Proposition 2.1.1. *Let $(E,\mathcal{E})$ and $(\tilde{E},\tilde{\mathcal{E}})$ be measurable spaces. If (v_t) is a stationary ergodic sequence of E–valued random elements and $f : (E^{\mathbb{N}},\mathcal{E}^{\mathbb{N}}) \to (\tilde{E},\tilde{\mathcal{E}})$ is measurable, then the sequence $(\tilde{v}_t)$ defined by*

$$\tilde{v}_t = f(v_t, v_{t-1},\dots), \qquad t \in \mathbb{Z}, \tag{2.1}$$

is stationary ergodic.

Proof. We demonstrate this result for the sake of completeness. Note that the measurable space $(E^{\mathbb{N}},\mathcal{E}^{\mathbb{N}})$ is defined analogously to $(E^{\mathbb{Z}},\mathcal{E}^{\mathbb{Z}})$. We follow the lines of proof of Proposition 4.3 in Section 1.4 of Krengel [77]. For stationarity we have to demonstrate that $(\tilde{v}_t) \overset{\mathrm{d}}{=} (\tilde{v}_{t+k})$ for every fixed $k \in \mathbb{Z}$. Let $\tilde{\tau}$ denote the backshift operator on $\tilde{E}^{\mathbb{Z}}$ and note that the special form (2.1) of f implies

$$f \circ \tau^k = \tilde{\tau}^k \circ f. \tag{2.2}$$

From this we conclude $\{(v_{t+k}) \in f^{-1}(\tilde{C})\} = \{(\tilde{v}_{t+k}) \in \tilde{C}\}$ for any $\tilde{C} \in \tilde{\mathcal{E}}^{\mathbb{Z}}$, and hence

$$\mathbb{P}((\tilde{v}_t) \in \tilde{C}) = \mathbb{P}((v_t) \in f^{-1}(\tilde{C})) = \mathbb{P}((v_{t+k}) \in f^{-1}(\tilde{C})) = \mathbb{P}((\tilde{v}_{t+k}) \in \tilde{C})$$

by stationarity of (v_t). Thus the stationarity of $(\tilde{v}_t)$ is shown. For proving the ergodicity of $(\tilde{v}_t)$, we assume that $\tilde{C} \in \tilde{\mathcal{E}}^{\mathbb{Z}}$ is an invariant set. We claim that $f^{-1}(\tilde{C})$ is invariant under the backshift operator τ on $E^{\mathbb{Z}}$. Indeed, by virtue of $f \circ \tau = \tilde{\tau} \circ f$ we have

$$f^{-1}(\tilde{C}) = f^{-1}(\tilde{\tau}^{-1}(\tilde{C})) = (\tilde{\tau} \circ f)^{-1}(\tilde{C}) = (f \circ \tau)^{-1}(\tilde{C}) = \tau^{-1}(f^{-1}(\tilde{C})).$$

Because $f^{-1}(\tilde{C})$ is invariant for (v_t), it is trivial and $\mathbb{P}((v_t) \in f^{-1}(\tilde{C}))$ either 0 or 1. The proof is completed by noting $\mathbb{P}((\tilde{v}_t) \in \tilde{C}) = \mathbb{P}((v_t) \in f^{-1}(\tilde{C}))$. □

Remark 2.1.2. An analogous statement is true for one–sided stationary ergodic sequences $(v_t)_{t\in\mathbb{N}}$: if $f : (E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}) \to (\tilde{E}, \tilde{\mathcal{E}})$ is measurable, then $(\tilde{v}_t)_{t\in\mathbb{N}} = (f(v_t, v_{t+1}, \ldots))_{t\in\mathbb{N}}$ is stationary ergodic; see Krengel [77]. □

The following corollary is particularly useful for showing stationarity and ergodicity of time series models which are defined through some limit relation.

Corollary 2.1.3. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and let $(E, \mathcal{E})$ be a measurable space and $\tilde{E}$ a complete and separable metric space, which we endow with its Borel σ–field $\tilde{\mathcal{E}} = \mathcal{B}(\tilde{E})$. Assume that (v_t) is a stationary ergodic sequence of E–valued random elements. Let $f_m : (E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}) \to (\tilde{E}, \tilde{\mathcal{E}})$, $m \in \mathbb{N}$, be measurable maps such that for some $t_0 \in \mathbb{Z}$ the sequence $(f_m(v_{t_0}, v_{t_0-1}, \ldots))_{m\in\mathbb{N}}$ converges in $\tilde{E}$* a.s. *Then there exists a measurable map $f : (E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}) \to (\tilde{E}, \tilde{\mathcal{E}})$ such that for all $t \in \mathbb{Z}$,*

$$\tilde{v}_t = \lim_{m\to\infty} f_m(v_t, v_{t-1}, \ldots) = f(v_t, v_{t-1}, \ldots) \quad \text{a.s.},$$

and the sequence $(\tilde{v}_t)$ is stationary ergodic.

Proof. Although the content of the corollary seems quite elementary, we are not aware of any reference.

Define the subset

$$E_0 = \left\{ (e_t) \in E^{\mathbb{N}} \;\middle|\; \lim_{m\to\infty} f_m((e_t)) \text{ exists} \right\}$$

and the function $f : (E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}) \to (\tilde{E}, \tilde{\mathcal{E}})$ by

$$f((e_t)) = \begin{cases} \lim_{m\to\infty} f_m((e_t)), & (e_t) \in E_0, \\ \tilde{e}_0, & (e_t) \notin E_0, \end{cases}$$

where $\tilde{e}_0 \in \tilde{E}$ is an arbitrary fixed element. The proof is completed if we can show that f is measurable. Because by construction of f we have that $f_m(v_t, v_{t-1}, \ldots) \to f(v_t, v_{t-1}, \ldots)$ as $m \to \infty$ if and only if $(v_t, v_{t-1} \ldots) \in E_0$. Since we assume that $f_m(v_{t_0}, v_{t_0-1}, \ldots)$ converges almost surely,

$$\mathbb{P}((v_{t_0}, v_{t_0-1}, \ldots) \in E_0) = 1. \tag{2.3}$$

Because of the stationarity of the sequence (v_t), relation (2.3) implies that $\mathbb{P}((v_t, v_{t-1}, \ldots) \in E_0) = 1$ for every $t \in \mathbb{Z}$, and consequently we have $\lim_{m\to\infty} f_m(v_t, v_{t-1}, \ldots) = f(v_t, v_{t-1}, \ldots)$ a.s. for *every* $t \in \mathbb{Z}$. The stationarity and ergodicity of the sequence $(\tilde{v}_t)$ then follow by an application of Proposition 2.1.1.

We now show that f is measurable. The ideas are based on Section 4.2 of Dudley [43]. Let us first prove that $E_0 \in \mathcal{E}^{\mathbb{N}}$. Since $\tilde{E}$ is complete, E_0 is equivalent to the set of elements $(e_t) \in E^{\mathbb{N}}$ with $(f_m((e_t)))_{m\in\mathbb{N}}$ being Cauchy in $\tilde{E}$, i.e., if d stands for the metric in $\tilde{E}$,

$$E_0 = \bigcap_{k\geq 1} \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} \Big\{ (e_t) \in E^{\mathbb{N}} \,\Big|\, \mathrm{d}(f_m((e_t)), f_n((e_t))) < 1/k \Big\}.$$

At this stage the separability assumption has to be exploited because one cannot directly conclude that the sets $\{ (e_t) \in E^{\mathbb{N}} \,|\, \mathrm{d}(f_m((e_t)), f_n((e_t))) < 1/k \}$ are measurable. There is a countable subset $M = \{\tilde{e}_1, \tilde{e}_2, \ldots\} \subset \tilde{E}$ whose closure is equal to $\tilde{E}$. We define $h_m : (E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}) \to (\tilde{E}, \tilde{\mathcal{E}})$ by

$$h_m((e_t)) = \mathrm{argmin}_{\{\tilde{e}_i \,|\, i=1,\ldots,m\}} \mathrm{d}(f_m((e_t)), \tilde{e}_i),$$

where by convention the minimizer is the element $\tilde{e}_i$ with lowest index i if two or more points are equally close to $f_m((e_t))$. Note that the range of h_m is included in the finite set $\{\tilde{e}_1, \ldots, \tilde{e}_m\}$. It is standard to verify that h_m is measurable and that $(f_m((e_t)))_{m\in\mathbb{N}}$ converges if and only if $(h_m((e_t)))_{m\in\mathbb{N}}$ converges. Thus we have the representation

$$E_0 = \bigcap_{k\geq 1} \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} \Big\{ (e_t) \in E^{\mathbb{N}} \,\Big|\, \mathrm{d}(h_m((e_t)), h_n((e_t))) < 1/k \Big\}.$$

Since h_m and h_n are measurable and have a *finite* range, one can show that the sets

$$\Big\{ (e_t) \in E^{\mathbb{N}} \,\Big|\, \mathrm{d}(h_m((e_t)), h_n((e_t))) < 1/k \Big\}$$

are measurable for all m, n and k, and consequently E_0 is measurable, i.e., $E_0 \in E^{\mathbb{N}}$. From this we can deduce that the functions

$$\tilde{f}_m((e_t)) = \begin{cases} f_m((e_t)), & (e_t) \in E_0, \\ \tilde{e}_0, & (e_t) \notin E_0 \end{cases}$$

are measurable for every $m \in \mathbb{N}$. Also note that $\tilde{f}_m \to f$ as $m \to \infty$ on $E^{\mathbb{N}}$. Since a sequence of measurable functions taking values in a metric space and converging everywhere has a measurable limit, see e.g. Theorem 4.2.2 in Dudley [43], f is measurable. This completes the proof of the corollary. □

It is now an easy matter to establish the stationarity and ergodicity of causal linear processes.

Example 2.1.4. Let (X_t) be a causal linear process, i.e., of the form

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \qquad t \in \mathbb{Z},$$

where $(\psi_j)_{j\in\mathbb{N}}$ is an absolutely summable sequence with $\psi_0 = 1$ and (Z_t) is an iid sequence with mean zero. Then, since (Z_t) is iid (and hence stationary ergodic by Kolmogorov's 0–1 law) and $X_t = \lim_{m\to\infty} \sum_{j=0}^{m} \psi_j Z_{t-j}$ a.s., the process (X_t) is stationary ergodic as well by Corollary 2.1.3. We remark that $\mathbb{E}\big(\sum_{j=0}^{m} |\psi_j|\,|Z_{t-j}|\big) < \infty$ by the monotone convergence theorem, which shows that $\sum_{j=0}^{m} |\psi_j|\,|Z_{t-j}| < \infty$ a.s.. □

Remark 2.1.5. Corollary 2.1.3 in its present formulation does not allow one to show the stationarity and ergodicity of *non*–causal linear processes, i.e., processes of the form

$$X_t = \sum_{j\in\mathbb{Z}} \psi_j Z_{t-j}, \qquad t \in \mathbb{Z},$$

where $\sum_{j\in\mathbb{Z}} |\psi_j| < \infty$ and (Z_t) is iid with mean zero. However, recalling that (2.2) was the crucial property used in the proof of Proposition 2.1.1, one may generalize Proposition 2.1.1 and Corollary 2.1.3 by merely requiring that the maps f and f_m obey the relation (2.2). The generalized results will then enable one to verify the stationarity and ergodicity of non–causal linear processes. We omit details. □

2.2 Uniform Convergence via the Ergodic Theorem

We begin by recalling Birkhoff's ergodic theorem [15] for *real*–valued sequences of random variables: if the sequence (X_t) or $(X_t)_{t\in\mathbb{N}}$, respectively, is stationary ergodic with $\mathbb{E}|X_0| < \infty$, then

$$\frac{1}{n}\sum_{t=1}^{n} X_t \xrightarrow{\text{a.s.}} \mathbb{E}X_0, \qquad n \to \infty.$$

Note also that the ergodic theorem is a generalization of the strong law of large numbers for iid sequences to dependent sequences. A proof of the ergodic theorem is contained in most advanced textbooks on probability theory.

In proofs for establishing consistency and asymptotic normality of M–estimators, one often has to show that the uniform strong law of large numbers is valid. To be more specific, let us assume that $K \subset \mathbb{R}^d$ is a compact set. Then $\mathbb{C}(K, \mathbb{R}^{d'})$ denotes the space of continuous $\mathbb{R}^{d'}$–valued functions on K and, for short, $\mathbb{C}(K, \mathbb{R}) = \mathbb{C}(K)$. We equip $\mathbb{C}(K, \mathbb{R}^{d'})$ with the supremum norm

$$\|v\|_K = \sup_{s \in K} |v(s)|, \qquad v \in \mathbb{C}(K, \mathbb{R}^{d'}),$$

where $|v(s)| = (v_1^2(s) + \cdots + v_{d'}^2(s))^{1/2}$ is the Euclidean norm of the vector $v(s) = (v_1(s), \ldots, v_{d'}(s))^T$. It is well–known that $\mathbb{C}(K, \mathbb{R}^{d'})$ is a separable Banach space, see e.g. Kufner et al. [78]. For enabling probability theory, we endow $\mathbb{C}(K, \mathbb{R}^{d'})$ with the Borel σ–field generated by the open sets in $\mathbb{C}(K, \mathbb{R}^{d'})$. Let us assume that we are given a stationary ergodic sequence of random elements (v_t)) (or $(v_t)_{t \in \mathbb{N}}$) with values in $\mathbb{C}(K, \mathbb{R}^{d'})$. In this context we say that the uniform strong law of large numbers holds, if in $\mathbb{C}(K, \mathbb{R}^{d'})$,

$$w_n = \frac{1}{n} \sum_{t=1}^{n} v_t \xrightarrow{\text{a.s.}} v, \qquad n \to \infty, \tag{2.4}$$

where the function v is defined pointwise by

$$v(s) = \mathbb{E}[v_0(s)], \qquad s \in K. \tag{2.5}$$

Note that (2.4) entails that $v \in \mathbb{C}(K, \mathbb{R}^{d'})$ a.s. Many authors establish the validity of the uniform strong law of large numbers in two steps. First they show pointwise convergence, i.e.,

$$w_n(s) = \frac{1}{n} \sum_{t=1}^{n} v_t(s) \xrightarrow{\text{a.s.}} v(s), \qquad n \to \infty,$$

for every *fixed* $s \in K$, e.g. by applying the ergodic theorem to the *real*–valued ergodic sequence $(v_t(s))$. In a second step, uniformity is proved, i.e.,

$$\|w_n - v\|_K \xrightarrow{\text{a.s.}} 0, \qquad n \to \infty.$$

Recall that a sequence of functions converges in $\mathbb{C}(K, \mathbb{R}^{d'})$ if and only if it converges pointwise and constitutes an equicontinuous family; see e.g. Lang [80]. Therefore the almost sure uniform convergence of $(w_n)_{n \geq 1}$ is equivalent to pointwise convergence and equicontinuity of the family $\{w_n \mid n \geq 1\}$ with probability one. In the literature there exist various conditions which imply equicontinuity, see e.g. Andrews [2]. These conditions often involve certain Lipschitz or Hölder conditions. For example the condition

$$\sup_{n \geq 1} \left\{ \sup_{s,u \in K, s \neq u} \frac{|w_n(s) - w_n(u)|}{|s - u|} \right\} < \infty \quad \text{a.s.}$$

implies the almost sure equicontinuity of $\{w_n \mid n \geq 1\}$. If one assumes that K is connected and v_t is continuously differentiable with probability 1, then the latter condition is implied by $\mathbb{E}\|v_0'\|_K < \infty$, as follows by an application of the mean value theorem and the ergodic theorem (note that (v_t') is stationary ergodic). This is actually the kind of condition which is verified in the articles of Lee and Hansen [84], Lumsdaine [90] and Berkes et al. [8]. We suggest a simpler method, which avoids the calculation of the first derivatives. It is based on the ergodic theorem for $\mathbb{C}(K, \mathbb{R}^{d'})$–valued sequences of random variables and requires that the stationary sequence (v_t) is ergodic and has a bounded expected norm.

Theorem 2.2.1. *Let (v_t) or $(v_t)_{t\in\mathbb{N}}$, respectively, be a stationary ergodic sequence of random elements with values in $\mathbb{C}(K, \mathbb{R}^{d'})$. Then the uniform strong law of large numbers is implied by*

$$\mathbb{E}\|v_0\|_K < \infty.$$

Hence the ergodicity of (v_t) together with $\mathbb{E}\|v_0\|_K < \infty$ implies the a.s. equicontinuity of $\{w_n \mid n \geq 1\}$. Theorem 2.2.1 goes back to Ranga Rao [115] and is e.g. contained in Parthasarathy [110]. In the subsequent sections we explain the background of Theorem 2.2.1 and provide a proof in the more general setting of a separable Banach space.

Exploiting the work of Wald [131] and Le Cam [81] on the consistency of the maximum likelihood estimator for *iid* data (see also Ferguson [51]), it is also possible to provide a proof of Theorem 2.2.1 merely based on the ergodic theorem for *real*–valued sequences together with certain topological arguments; we omit the details.

2.2.1 Bochner Expectation

We assume that B denotes a separable real Banach space with norm $\|\cdot\|$, endowed with its Borel σ–field $\mathcal{B}$ generated by the open sets with respect to $\|\cdot\|$. The separability of B assures that any linear combination of B–valued random elements is again measurable (and therefore a random element); see Ledoux and Talagrand [83]. However, also the expectation of B–valued random elements needs to be defined. If B has finite dimension d, the space B is isomorphic to $\mathbb{R}^d$, and taking the expectation is essentially coordinate–wise (ordinary) Lebesgue expectation. A general infinite–dimensional space B however calls for an extension of the Lebesgue expectation. This task was accomplished in 1933 by Bochner [17]. Since B is assumed to be separable, we may use Lemma V-2-4 and Proposition V-2-5 in Neveu [107] for introducing the Bochner expectation $\mathbb{E}_{\mathrm{Bo}}$. It is defined for B–valued random elements Y with $\mathbb{E}\|Y\| < \infty$. Note that $\|Y\|$ is automatically an ordinary nonnegative random variable because taking the norm is a continuous and therefore measurable transformation. Lemma V-2-4 in Neveu [107] says that such a Y may be approximated by simple random elements, i.e., by random elements of the form

$$Y_m = \sum_{i=1}^{k_m} c_{m,i}\, \mathbf{1}_{A_{m,i}},$$

where $c_{m,i} \in B$ are deterministic and $A_{m,i} \in \mathcal{A}$ constitute a disjoint decomposition of Ω, such that

(a) for every $\omega \in \Omega$: $Y(\omega) = \lim_{m\to\infty} Y_m(\omega)$,

(b) for every $\omega \in \Omega$ and $m \in \mathbb{N}$: $\|Y_m(\omega)\| \le \|Y(\omega)\|$,

(c) $\lim_{m\to\infty} \mathbb{E}\|Y_m - Y\| = 0$.

This leads to a natural definition of the Bochner expectation $\mathbb{E}_{\mathrm{Bo}}Y$ with respect to the probability measure $\mathbb{P}$:

1. For a simple random element $Y = \sum_{i=1}^{k} c_i \mathbf{1}_{A_i}$ we define the Bochner expectation by

$$\mathbb{E}_{\mathrm{Bo}}Y = \sum_{i=1}^{k} c_i\, \mathbb{P}(A_i).$$

2. A general random element Y is approximated by a sequence of simple random elements $(Y_m)_{m\in\mathbb{N}}$ fulfilling $\lim_{m\to\infty} \mathbb{E}\|Y_m - Y\| = 0$. The existence of such a sequence is guaranteed by Lemma V-2-4 in Neveu [107]. It is easy to show that $\|\mathbb{E}_{\mathrm{Bo}}Z\| \le \mathbb{E}\|Z\|$ for any simple random element Z. This inequality and the facts that linear combinations of simple random elements are again simple and that the operator $\mathbb{E}_{\mathrm{Bo}}$ is linear in the space of simple random elements. imply

$$\|\mathbb{E}_{\mathrm{Bo}}Y_{m+j} - \mathbb{E}_{\mathrm{Bo}}Y_m\| \le \mathbb{E}\|Y_{m+j} - Y_m\| \le \mathbb{E}\|Y_{m+j} - Y\| + \mathbb{E}\|Y_m - Y\|.$$

Since $\mathbb{E}\|Y_m - Y\| \to 0$, this shows that $(\mathbb{E}_{\mathrm{Bo}}Y_m)_{m\in\mathbb{N}}$ forms a Cauchy sequence in B and hence has a limit. We define the Bochner expectation by

$$\mathbb{E}_{\mathrm{Bo}}Y = \lim_{m\to\infty} \mathbb{E}_{\mathrm{Bo}}Y_m.$$

The Bochner expectation is well–defined since the definition of $\mathbb{E}_{\mathrm{Bo}}Y$ is irrespective of the approximating sequence. Indeed, if $(\tilde{Y}_m)_{m\in\mathbb{N}}$ is another sequence fulfilling $\lim_{m\to\infty} \mathbb{E}\|\tilde{Y}_m - Y\| = 0$, then by

$$\begin{aligned}\|\mathbb{E}_{\mathrm{Bo}}Y_m - \mathbb{E}_{\mathrm{Bo}}\tilde{Y}_m\| &= \|\mathbb{E}_{\mathrm{Bo}}(Y_m - \tilde{Y}_m)\| \le \mathbb{E}\|Y_m - \tilde{Y}_m\| \\ &\le \mathbb{E}\|Y_m - Y\| + \mathbb{E}\|Y - \tilde{Y}_m\| \to 0\end{aligned}$$

we may conclude that $\lim_{m\to\infty} \mathbb{E}_{\mathrm{Bo}}Y_m = \lim_{n\to\infty} \mathbb{E}_{\mathrm{Bo}}\tilde{Y}_m$.

3. It is immediate that the Bochner expectation is linear and decreases the norm, that means for any random elements Y_1 and Y_2 with $\mathbb{E}||Y_i|| < \infty$, $i = 1, 2$, and $a_1, a_2 \in \mathbb{R}$

$$\mathbb{E}_{\text{Bo}}(a_1 Y_1 + a_2 Y_2) = a_1 \mathbb{E}_{\text{Bo}} Y_1 + a_2 \mathbb{E}_{\text{Bo}} Y_2,$$

$$||\mathbb{E}_{\text{Bo}} Y_1|| \leq \mathbb{E}||Y_1||.$$

4. Because the Bochner expectation of simple real–valued random variables coincides with the Lebesgue expectation, the Bochner expectation of real–valued random variables is equivalent to the Lebesgue expectation; hence Bochner expectation can be regarded as a generalization of Lebesgue expectation.

Remark 2.2.2. The Bochner expectation may also be defined for functions taking values in a *non*–separable real Banach space, but then Lemma V-2-4 in Neveu [107] cannot be applied and only those Y's are integrable, for which there exists an approximating sequence $(Y_m)_{m\in\mathbb{N}}$ of simple random elements such that $\lim_{m\to\infty} Y_m = Y$ a.s. and $\lim_{m\to\infty} \mathbb{E}||Y_m - Y|| = 0$. This also means that $\mathbb{E}||Y|| < \infty$ is in general not sufficient anymore for Bochner integrability; cf. Yosida [134]. □

Later it will be essential that we can make use of the change of variables formula. Fortunately it carries over to Bochner expectation. This is formalized in the following proposition.

Proposition 2.2.3. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $(E, \mathcal{E})$ a measurable space and $h : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ a measurable map. Let B be a separable real Banach space and Y be a B–valued random element on E. Denote by $h \circ \mathbb{P}$ the measure on $(E, \mathcal{E})$ induced by h. Then the following is true:*

$$\mathbb{E}_{h\circ\mathbb{P}}||Y|| < \infty \iff \mathbb{E}_{\mathbb{P}}||Y \circ h|| < \infty,$$

Hence Y is Bochner integrable with respect to $h \circ \mathbb{P}$ if and only if $Y \circ h$ is Bochner integrable with respect to $\mathbb{P}$. In that case, we have the change of variables formula

$$\mathbb{E}_{\mathbb{P},\text{Bo}}(Y \circ h) = \mathbb{E}_{h\circ\mathbb{P},\text{Bo}}(Y).$$

Proof. The first relation follows from the observation that $||Y \circ h|| = ||Y|| \circ h$ and from the change of variables formula for real–valued random variables. It is easy to verify that the change of variables formula is valid for simple random elements. Therefore let Y be a general random element which is Bochner integrable with respect to the measure $h \circ \mathbb{P}$. Then there exists a sequence of simple random elements $(Y_m)_{m\in\mathbb{N}}$ so that $\lim_{m\to\infty} \mathbb{E}_{h\circ\mathbb{P}}||Y_m - Y|| = 0$. Since by the change of variables formula for real–valued random variables,

$$\mathbb{E}_{h\circ\mathbb{P}}||Y_m - Y|| = \mathbb{E}_{\mathbb{P}}(||Y_m - Y|| \circ h) = \mathbb{E}_{\mathbb{P}}||Y_m \circ h - Y \circ h||,$$

the simple random elements $Y_m \circ h$ also fulfill $\lim_{m\to\infty} \mathbb{E}_{\mathbb{P}}\|Y_m \circ h - Y \circ h\| = 0$. Therefore by definition of the Bochner expectation and the change of variables formula for simple random elements,

$$\mathbb{E}_{\mathbb{P},\mathrm{Bo}}(Y \circ h) = \lim_{m\to\infty} \mathbb{E}_{\mathbb{P},\mathrm{Bo}}(Y_m \circ h) = \lim_{m\to\infty} \mathbb{E}_{h\circ\mathbb{P},\mathrm{Bo}}(Y_m) = \mathbb{E}_{h\circ\mathbb{P},\mathrm{Bo}}(Y),$$

which concludes the proof. □

In the following proposition we have a closer look at the Bochner expectation of $\mathbb{C}(K, \mathbb{R}^{d'})$–valued random elements. Recall that $\mathbb{C}(K, \mathbb{R}^{d'})$ is a separable real Banach space. There is a tight relationship between Bochner expectation and pointwise Lebesgue expectation. As usual we use the supremum norm in $\mathbb{C}(K, \mathbb{R}^{d'})$ and impose the Borel σ–field generated by the open sets.

Proposition 2.2.4. *A $\mathbb{C}(K, \mathbb{R}^{d'})$–valued random element Y is Bochner integrable if and only if*

$$\mathbb{E}\|Y\|_K < \infty,$$

and in that case $\mathbb{E}_{\mathrm{Bo}} Y$ is characterized by

$$\mathbb{E}_{\mathrm{Bo}}[Y](s) = \mathbb{E}[Y(s)], \qquad s \in K.$$

Proof. The criterion for Bochner integrability is immediate. Therefore we only have to show the validity of the characterization. Let $s \in K$ be fixed. First note that $Y(s)$ is measurable because it results from a continuous transform of Y. The relation $\mathbb{E}_{\mathrm{Bo}}[Y](s) = \mathbb{E}[Y(s)]$, $s \in K$, clearly holds for simple random elements Y. For general Y we take the approximating sequence $(Y_m)_{m\in\mathbb{N}}$ of simple random elements constructed in Lemma V-2-4 of Neveu [107], which yields $\mathbb{E}_{\mathrm{Bo}} Y = \lim_{m\to\infty} \mathbb{E}_{\mathrm{Bo}} Y_m$. Because uniform convergence of functions trivially implies pointwise convergence, also

$$\mathbb{E}_{\mathrm{Bo}}[Y](s) = \lim_{m\to\infty} \mathbb{E}_{\mathrm{Bo}}[Y_m](s) = \lim_{m\to\infty} \mathbb{E}[Y_m(s)].$$

Since $\lim_{m\to\infty} Y_m(s) = Y(s)$ and $|Y_m(s)| \le \|Y_m\|_K \le \|Y\|_K$ for every $\omega \in \Omega$, as a consequence of the Lebesgue dominated convergence theorem we obtain $\lim_{m\to\infty} \mathbb{E}[Y_m(s)] = \mathbb{E}[Y(s)]$ a.s. This proves the assertion. □

2.2.2 The Ergodic Theorem for Sequences of B–valued Random Elements

The following theorem is implied by Theorem 2.1 in Section 4.2 of Krengel [77].

Theorem 2.2.5. *Let $(B, \mathcal{B})$ be a separable Banach space and (v_t) or $(v_t)_{t\in\mathbb{N}}$, respectively, a stationary ergodic sequence of B–valued random elements such that $\mathbb{E}\|v_0\| < \infty$. Then*

$$\frac{1}{n}\sum_{t=1}^{n} v_t \xrightarrow{\text{a.s.}} \mathbb{E}_{\mathrm{Bo}} v_0, \qquad n \to \infty. \tag{2.6}$$

Proof. We only provide the proof in the case of doubly–infinite stationary ergodic sequences. Let $\tilde{\mathbb{P}}$ denote the probability measure on $(B^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}})$ induced by (v_t) and let $f : (B^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}) \to (B, \mathcal{B})$ be the measurable map defined by $f((e_t)) = e_1$. Recall that τ is the backshift operator. Then we have the relation

$$\frac{1}{n}\sum_{t=1}^{n} v_t = \frac{1}{n}\sum_{k=0}^{n-1} f(\tau^k((v_t))).$$

The probability space $(B^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \tilde{\mathbb{P}})$ together with f and the measure preserving transformation τ fulfill the conditions of the Banach space–valued version of the ergodic theorem, see e.g. Theorem 2.1 in Section 4.2 of Krengel [77]; note in particular that f is Bochner integrable with respect to $\tilde{\mathbb{P}}$ since by the change of variables formula (Proposition 2.2.3),

$$\mathbb{E}_{\tilde{\mathbb{P}}} \|f\| = \mathbb{E}\|v_0\| < \infty.$$

Therefore for $\tilde{\mathbb{P}}$–almost every $(e_t) \in B^{\mathbb{Z}}$,

$$\frac{1}{n}\sum_{k=0}^{n-1} f(\tau^k((e_t))) \to \mathbb{E}_{\tilde{\mathbb{P}},\mathrm{Bo}}(f), \qquad n \to \infty.$$

This is equivalent to saying that $\mathbb{P}$–almost surely,

$$\frac{1}{n}\sum_{t=1}^{n} v_t \to \mathbb{E}_{\tilde{\mathbb{P}},\mathrm{Bo}}(f), \qquad n \to \infty.$$

Now another application of the change of variables formula for the Bochner expectation yields $\mathbb{E}_{\tilde{\mathbb{P}},\mathrm{Bo}}(f) = \mathbb{E}_{\mathrm{Bo}} v_0$ and completes the proof. □

To finish this section we mention that the proof of Theorem 2.2.1 is a consequence of Proposition 2.2.4 and Theorem 2.2.5, where we choose $B = \mathbb{C}(K, \mathbb{R}^{d'})$.

2.3 Matrix Norms

In this monograph we use two different matrix norms, as the case may be; since all matrix norms are equivalent, the particular choice of a norm is (mostly) irrelevant from a mathematical point of view. Recall that the Frobenius norm of a matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{d\times d}$, is defined by

$$\|\mathbf{A}\| = \Big(\sum_{i,j=1}^{d} a_{ij}^2\Big)^{1/2}. \tag{2.7}$$

The following inequality is valid for the Frobenius norm: if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ are column vectors and $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$, then the Frobenius norm of the matrix $\mathbf{x}\mathbf{y}^T$ is bounded by

$$\|\mathbf{x}\mathbf{y}^T\| \leq |\mathbf{x}|\,|\mathbf{y}|. \tag{2.8}$$

Occasionally we use the matrix operator norm with respect to the Euclidean norm instead, i.e.,

$$\|\mathbf{A}\|_{\mathrm{op}} = \sup_{\mathbf{x}\neq \mathbf{0}} \frac{|\mathbf{A}\mathbf{x}|}{|\mathbf{x}|}. \tag{2.9}$$

Finally we define the norm of a continuous matrix–valued function $\mathbf{A}$ on a compact set $K \subset \mathbb{R}^d$, i.e $\mathbf{A} \in \mathbb{C}(K, \mathbb{R}^{d' \times d'})$, by

$$\|\mathbf{A}\|_K = \sup_{\mathbf{s}\in K} \|\mathbf{A}(\mathbf{s})\|. \tag{2.10}$$

Of course Theorem 2.2.1 carries over to sequences of matrices: a stationary ergodic sequence of random elements $(\mathbf{A}_t)$ with values in $\mathbb{C}(K, \mathbb{R}^{d' \times d'})$ and with $\mathbb{E}\|\mathbf{A}_0\|_K < \infty$ fulfills

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{A}_t \xrightarrow{\text{a.s.}} \mathbf{M}_0, \qquad n \to \infty,$$

where $\mathbf{M}_0(\mathbf{s}) = \mathbb{E}[\mathbf{A}_0(\mathbf{s})]$, $\mathbf{s} \in K$; note that the expected value of a random matrix is defined element-wise. Moreover, inequality (2.8) carries over: for any two $\mathbf{x}, \mathbf{y} \in \mathbb{C}(K, \mathbb{R}^{d'})$,

$$\|\mathbf{x}\mathbf{y}^T\|_K \leq \|\mathbf{x}\|_K\,\|\mathbf{y}\|_K. \tag{2.11}$$

2.4 Weak Convergence in $\mathbb{C}(K, \mathbb{R}^{d'})$

While Section 2.2 dealt with the almost sure convergence of $\mathbb{C}(K, \mathbb{R}^{d'})$–valued sequences, this section summarizes a number of useful results concerning weak convergence in $\mathbb{C}(K, \mathbb{R}^{d'})$, which will be applied in Chapter 8. We assume that the reader is familiar with the definition of weak convergence in separable metric spaces; otherwise we refer to Billingsley [12].

The first well–known result gives sufficient and necessary conditions for weak convergence in $\mathbb{C}(K, \mathbb{R}^{d'})$.

Theorem 2.4.1. *Let $v, (v_n)_{n\in\mathbb{N}}$ be random elements with values in $\mathbb{C}(K, \mathbb{R}^{d'})$. Then the following two statements are equivalent:*

(i)

$$v_n \xrightarrow{\text{d}} v, \qquad n \to \infty.$$

(ii) (a) *The finite-dimensional distributions converge, i.e., for any* $s_1,\dots,s_m \in K$,

$$(v_n(s_1),\dots,v_n(s_m)) \xrightarrow{\mathrm{d}} (v(s_1),\dots,v(s_m)), \qquad n\to\infty.$$

(b) *For every* $\epsilon > 0$,

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} \mathbb{P}\Big(\sup_{|s-t|\le\delta} |v_n(s)-v_n(t)| > \epsilon\Big) = 0.$$

For a proof consult e.g. Billingsley [12], Theorem 7.3. The following corollary is an important implication of Theorem 2.4.1 and useful for the analysis of M–estimators.

Corollary 2.4.2. *Let* $(v_n)_{n\in\mathbb{N}}$ *be a* $\mathbb{C}(K,\mathbb{R}^{d'})$*–valued sequence of random elements having weak limit* v *in* $\mathbb{C}(K,\mathbb{R}^{d'})$*. Then for any sequence of random variables* $(T_n)_{n\in\mathbb{N}}\subset K$ *which converges to* $s_0\in K$ *in probability,*

$$v_n(T_n) \xrightarrow{\mathrm{d}} v(s_0), \qquad n\to\infty.$$

Proof. Since $v_n(s_0)\xrightarrow{\mathrm{d}} v(s_0)$, it suffices to show that $|v_n(T_n)-v_n(s_0)| \xrightarrow{\mathbb{P}} 0$ as $n\to\infty$. Let $\epsilon>0$ be fixed. Then for any $\delta>0$,

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\left(|v_n(T_n)-v_n(s_0)|>\epsilon\right)\\
&\le \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{|s-t|\le\delta}|v_n(s)-v_n(t)|>\epsilon,\ |T_n-s_0|\le\delta\Big)\\
&\quad + \limsup_{n\to\infty} \mathbb{P}\left(|T_n-s_0|>\delta\right)\\
&\le \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{|s-t|\le\delta}|v_n(s)-v_n(t)|>\epsilon\Big).
\end{aligned}$$

By Theorem 2.4.1 the right–hand side of the latter inequality converges to zero as $\delta\downarrow 0$. □

The next theorem contains a statement about sequences with two indices. It is an adaptation of Theorem 3.2 in Billingsley [12].

Theorem 2.4.3. *Suppose that* $(u_{nm},u_n)_{n\in\mathbb{N}}$ *and* $(v_m)_{m\in\mathbb{N}}$ *are sequences of random elements taking values in* $\mathbb{C}(K,\mathbb{R}^{d'})\times\mathbb{C}(K,\mathbb{R}^{d'})$ *and* $\mathbb{C}(K,\mathbb{R}^{d'})$*, respectively. Assume* v *is a* $\mathbb{C}(K,\mathbb{R}^{d'})$*–valued random element. Then if*

(a) $u_{nm}\xrightarrow{\mathrm{d}} v_m$ *as* $n\to\infty$,

(b) $v_m\xrightarrow{\mathrm{d}} v$ *as* $m\to\infty$,

and if

(c) *for every* $\epsilon > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big(\|u_{nm} - u_n\|_K > \epsilon \Big) = 0,$$

then also $u_n \xrightarrow{\text{d}} v$ *as* $n \to \infty$.

2.5 Exponentially Fast Almost Sure Convergence

Let $T = \mathbb{N}$ or $T = \mathbb{Z}$. For sequences $(v_t)_{t\in\mathbb{Z}}$ we write (v_t) in abridged form. The following definition will facilitate the formulation of many of our results.

Definition. A sequence $(v_t)_{t\in T}$ of random elements with values in a normed vector space $(B, \|\cdot\|)$ is said to converge to zero exponentially fast almost surely as $t \to \infty$, henceforth denoted by $v_t \xrightarrow{\text{e.a.s.}} 0$ if there exists $\gamma > 1$ with $\gamma^t \|v_t\| \xrightarrow{\text{a.s.}} 0$. □

Various proofs in this monograph rest on the following proposition, which is about the absolute convergence of sums with stochastic weights decaying exponentially fast. The proposition also shows that exponentially fast decay towards zero is preserved under component–wise multiplication with a stationary sequence having a finite positive logarithmic moment. Let $\log^+(x) = \log x$ if $x > 1$ and 0 otherwise.

Proposition 2.5.1. *Let* $(\xi_t)_{t\in T}$ *be a sequence of real random variables with* $\xi_t \xrightarrow{\text{e.a.s.}} 0$ *and* $(v_t)_{t\in T}$ *a sequence of identically distributed random elements with values in a separable Banach space* $(B, \|\cdot\|)$. *If* $\mathbb{E}(\log^+ \|v_0\|) < \infty$, *then* $\sum_{t=0}^{\infty} \xi_t v_t$ *converges* a.s., *and one has* $\xi_n \sum_{t=0}^{n} v_t \xrightarrow{\text{e.a.s.}} 0$ *and* $\xi_n v_n \xrightarrow{\text{e.a.s.}} 0$ *as* $n \to \infty$.

The proof of the proposition is basically a consequence of the following standard result.

Lemma 2.5.2. *If* $(X_t)_{t\in\mathbb{N}}$ *is a sequence of identically distributed random variables with* $\mathbb{E}(\log^+ |X_0|)$, *then for any* $|\rho| < 1$ *one has that* $\sum_{t=0}^{\infty} \rho^t X_t$ *is absolutely convergent almost surely.*

Proof. Without loss of generality, $X_t \geq 0$ and $0 < \rho < 1$. Define events $A_t = \{X_t > \rho^{-t/2}\}$. Then due to the fact that the X_t's are identically distributed,

$$\sum_{t=1}^{\infty} \mathbb{P}(A_t) = \sum_{t=1}^{\infty} \mathbb{P}(X_t > \rho^{-t/2}) = \sum_{t=1}^{\infty} \mathbb{P}(\log^+ X_0 > -(t/2)\log\rho)$$
$$\leq -\frac{2}{\log\rho}\, \mathbb{E}(\log^+ X_0) < \infty,$$

where the last inequality is a consequence of $\mathbb{E}Y = \int_0^\infty \mathbb{P}(Y > t)\,\mathrm{d}t < \infty$ for any random variable $Y \geq 0$. Thus by the Borel–Cantelli lemma, with probability one A_t occurs only finitely many times, i.e., $\rho^t X_t \leq \rho^{t/2}$ for all but finitely many t's. Since $\sum_{t=0}^{\infty} \rho^{t/2} < \infty$, the lemma is proved. □

Proof of Proposition 2.5.1. Note that the separability of B is needed in order to guarantee that the partial sums $\sum_{t=0}^{n} \xi_t v_t$ are again measurable, see e.g. Ledoux and Talagrand [83]. Since $\xi_t \xrightarrow{\text{e.a.s.}} 0$, there is $\gamma > 1$ with $\gamma^t |\xi_t| \xrightarrow{\text{a.s.}} 0$. Hence there exists a random variable $C \geq 0$ with $|\xi_t| \leq C\gamma^{-t}$ a.s. for all $t \in \mathbb{N}$. The facts that $\gamma^{-1} < 1$ and $\mathbb{E}(\log^+ \|v_0\|) < \infty$ imply $\sum_{t=0}^{\infty} \gamma^{-t} \|v_t\| < \infty$ a.s. by Lemma 2.5.2. Therefore

$$\sum_{t=0}^{\infty} \|\xi_t v_t\| = \sum_{t=0}^{\infty} |\xi_t| \, \|v_t\| \leq C \sum_{t=0}^{\infty} \gamma^{-t} \|v_t\| < \infty \quad \text{a.s.}$$

Since B is complete, this implies that $\sum_{t=0}^{\infty} \xi_t v_t$ converges in B a.s. As to the second part of the lemma, choose $\tilde{\gamma} > 1$ small enough such that $\eta = \tilde{\gamma}/\gamma < 1$. Since $\eta^{1/2} < 1$ and $\mathbb{E}(\log^+ \|v_0\|) < \infty$ imply $\sum_{t=0}^{\infty} \eta^{t/2} \|v_t\| < \infty$ a.s., the estimate

$$\tilde{\gamma}^n \left\| \xi_n \sum_{t=0}^{n} v_t \right\| = \eta^n \gamma^n |\xi_n| \left\| \sum_{t=0}^{n} v_t \right\| \leq C\eta^{n/2} \sum_{t=0}^{n} \eta^{n/2} \|v_t\| \leq C\eta^{n/2} \sum_{t=0}^{\infty} \eta^{t/2} \|v_t\|$$

establishes $\xi_n \sum_{t=0}^{n} v_t \xrightarrow{\text{e.a.s.}} 0$ as $n \to \infty$ and $\xi_n v_n \xrightarrow{\text{e.a.s.}} 0$. This completes the proof. □

Note that $\mathbb{E}(\log^+ \|v_0\|) < \infty$ is implied by the existence of some power moment $\mathbb{E}\|v_0\|^q < \infty$, $q > 0$, because $\log^+ y = o(y^q)$ as $y \to \infty$. If two random elements have a finite expected positive logarithmic norm, then so do their sum and the product of their norms. This follows from the facts that $\log^+(y_1 + y_2) \leq \log^+(2\max(y_1 y_2)) \leq \log 2 + \log^+ y_1 + \log^+ y_2$ and $\log^+(y_1 y_2) \leq \log^+ y_1 + \log^+ y_2$ for all $y_1, y_2 \geq 0$. These properties, which are summarized in the following lemma, will be often used without any explicit reference.

Lemma 2.5.3. *Let v_1 and v_2 be two random elements taking values in a separable Banach space. Then the existence of a $q > 0$ such that $\mathbb{E}\|v_1\|^q < \infty$ implies $\mathbb{E}(\log^+ \|v_1\|) < \infty$. Moreover*

$$\mathbb{E}(\log^+ \|v_1 + v_2\|) \leq \log 2 + \mathbb{E}(\log^+ \|v_1\|) + \mathbb{E}(\log^+ \|v_2\|)$$

and

$$\mathbb{E}[\log^+(\|v_1\| \, \|v_2\|)] \leq \mathbb{E}(\log^+ \|v_1\|) + \mathbb{E}(\log^+ \|v_2\|).$$

We also need the following auxiliary result.

Lemma 2.5.4. *Let $(Y_t)_{t \in T}$ and $(\tilde{Y}_t)_{t \in T}$ be sequences with $|Y_t - \tilde{Y}_t| \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$. Then:*

(i) $|Y_t|^{1/2} - |\tilde{Y}_t|^{1/2} \xrightarrow{\text{e.a.s.}} 0$ *as* $t \to \infty$.

(ii) *If in addition* $(Y_t)_{t\in T}$ *is stationary with* $\mathbb{E}|Y_0| < \infty$*, then also* $|\exp(Y_t) - \exp(\tilde{Y}_t)| \xrightarrow{\text{e.a.s.}} 0$.

Proof. (i): Without loss of generality, $Y_t, \tilde{Y}_t \geq 0$. By assumption on $|Y_t - \tilde{Y}_t|$, there is a $\gamma > 1$ with $\gamma^t|Y_t - \tilde{Y}_t| \xrightarrow{\text{a.s.}} 0$. Consequently there exists t_0 (depending on the realization ω) such that for each $t \geq t_0$, either $\{\min(Y_t, \tilde{Y}_t) \leq \gamma^{-t}, \max(Y_t, \tilde{Y}_t) \leq \gamma^{-t/2}\}$ or $\{\min(Y_t, \tilde{Y}_t) > \gamma^{-t}\}$. Using this observation and by an application of the mean value theorem to the difference of square roots, we conclude

$$|(Y_t)^{1/2} - (\tilde{Y}_t)^{1/2}| \leq \max\left\{\gamma^{-t/2}, 2^{-1}\gamma^{t/2}|Y_t - \tilde{Y}_t|\right\}, \qquad t \geq t_0.$$

This shows that $\tilde{\gamma}^t|(Y_t)^{1/2} - (\tilde{Y}_t)^{1/2}| \xrightarrow{\text{a.s.}} 0$ if $1 < \tilde{\gamma} < \gamma^{1/2}$ and thus $|(Y_t)^{1/2} - (\tilde{Y}_t)^{1/2}| \xrightarrow{\text{e.a.s.}} 0$.

(ii): As soon as $|Y_t - \tilde{Y}_t| \leq 1$, by the mean value theorem

$$|\exp(Y_t) - \exp(\tilde{Y}_t)| = \exp(Y_t)\,|1 - \exp(\tilde{Y}_t - Y_t)| \leq \exp(1 + Y_t)\,|Y_t - \tilde{Y}_t|.$$

Now the claim follows from $|Y_t - \tilde{Y}_t| \xrightarrow{\text{e.a.s.}} 0$ together with $\mathbb{E}[\log^+\{\exp(1 + Y_0)\}] \leq 1 + \mathbb{E}|Y_0| < \infty$ and an application of Proposition 2.5.1. This concludes the proof. □

The following elementary result about random products is used in abundance in this monograph.

Lemma 2.5.5. *Let* $(Y_t)_{t\in T}$ *be a stationary ergodic sequence of random variables with* $\mathbb{E}(\log|Y_0|) < 0$*. Then*

$$\prod_{t=0}^{n} Y_t \quad \xrightarrow{\text{e.a.s.}} \quad 0, \qquad n \to \infty.$$

If, in addition, $(Y_t)_{t\in T}$ *is an iid sequence with* $\mathbb{E}|Y_0|^q < \infty$ *for some* $q > 0$*, then there exist* $0 < \tilde{q} \leq q$ *and* $0 < \lambda < 1$ *such that*

$$\mathbb{E}\Big(\Big|\prod_{t=0}^{n-1} Y_t\Big|^{\tilde{q}}\Big) = \lambda^n, \qquad n \to \infty.$$

Proof. The first assertion follows from a straightforward application of the ergodic theorem to the logarithm of $\prod_{t=0}^{n}|Y_t|$ together with $\mathbb{E}(\log|Y_0|) < 0$. Concerning the second claim, observe that the map $s \mapsto \mathbb{E}|Y_0|^s$ on $[0, q]$ has first derivative equal to $\mathbb{E}(\log|Y_0|) < 0$ at $s = 0$. This shows the existence of a $\tilde{q} \in (0, q]$ with $\lambda := \mathbb{E}|Y_0|^{\tilde{q}} < 1$, and since $(Y_t)_{t\in T}$ is iid, $\mathbb{E}(|\prod_{t=0}^{n-1} Y_t|^{\tilde{q}}) = (\mathbb{E}|Y_0|^{\tilde{q}})^n = \lambda^n$. This concludes the proof of the lemma. □

2.6 Stochastic Recurrence Equations

We have chosen to base our treatment of inference in conditionally heteroscedastic time series models on stochastic recurrence equations (SREs) because we believe that such an approach is well adapted to the problems one encounters. Stochastic recurrence equation (SRE) techniques play a prominent role in classical time series analysis, albeit this is often not explicitly expressed in the literature. One also has to be aware of the fact that every homogeneous Markov chain can be seen as a solution to a SRE, see e.g. Proposition 7.6 in Kallenberg [71]. We think that it is worthwhile to lay down a mathematically precise notion of a SRE. Before we start, we would like to mention the excellent survey article by Diaconis and Freedman [38] for an overview and other nice applications of SREs.

Although we merely work in Banach spaces, we present the notion of a stochastic recurrence equation (SRE) in the more general case of a complete separable metric space (i.e. Polish space), as it was e.g. formulated by Bougerol [20]. Hence, let (E, d) be a Polish space equipped with its Borel σ–field $\mathcal{B}(E)$. Recall that a map $\phi : E \to E$ is called Lipschitz if

$$\Lambda(\phi) := \sup_{x,y\in E, x\neq y} \frac{\mathrm{d}(\phi(x), \phi(y))}{\mathrm{d}(x,y)} . \tag{2.12}$$

is finite and called a contraction if $\Lambda(\phi) < 1$. Also note that Λ is submultiplicative, i.e., if ϕ and ψ are Lipschitz maps $E \to E$, then

$$\Lambda(\phi \circ \psi) \leq \Lambda(\phi)\,\Lambda(\psi). \tag{2.13}$$

We consider a process (ϕ_t) of random Lipschitz maps $E \to E$ with $\phi_t(x)$ being $\mathcal{B}(E)$–measurable for every fixed $x \in E$ and $t \in \mathbb{Z}$. In what follows, $T = \mathbb{N}$ or $T = \mathbb{Z}$. If for a stochastic process $(X_t)_{t\in T}$ with values in E,

$$X_{t+1} = \phi_t(X_t) \quad \text{a.s.}, \qquad t \in T, \tag{2.14}$$

we say that $(X_t)_{t\in T}$ obeys the SRE associated with (ϕ_t). Alternatively, $(X_t)_{t\in T}$ is referred to as a solution to the SRE (2.14).

We write $(\phi_t^{(n)})_{n\in\mathbb{N}}$ for the sequence of the n–fold iterations of past and present transformations, i.e.,

$$\phi_t^{(n)} = \begin{cases} \mathrm{Id}_E, & n = 0, \\ \phi_t \circ \phi_{t-1} \circ \cdots \circ \phi_{t-n+1}, & n \geq 1, \end{cases} \tag{2.15}$$

where Id_E is the identity map in E. The following theorem due to Bougerol [20] is a generalization of results by Letac [86] and can be considered as a stochastic version of Banach's fixed point theorem. It makes statements about the solutions of the SRE (2.14) under the assumption that (ϕ_t) is stationary ergodic and that ϕ_0 or a certain r–fold iterate $\phi_0^{(r)}$ is "contractive on average".

Theorem 2.6.1 (Theorem 3.1 of Bougerol [20]). *Let* (ϕ_t) *be a stationary ergodic sequence of Lipschitz maps from* E *into* E*. Suppose the following conditions hold:*

S.1 *There is* $y \in E$ *such that* $\mathbb{E}[\log^+ \mathrm{d}(\phi_0(y), y)] < \infty$.

S.2 $\mathbb{E}[\log^+ \Lambda(\phi_0)] < \infty$ *and for some integer* $r \geq 1$,

$$\mathbb{E}[\log \Lambda(\phi_0^{(r)})] = \mathbb{E}[\log \Lambda(\phi_0 \circ \cdots \circ \phi_{-r+1})] < 0. \tag{2.16}$$

Then the SRE (2.14) *admits a stationary ergodic solution* $(Y_t)_{t \in T}$*, which has the stochastic representation*

$$Y_t = \lim_{m \to \infty} \phi_{t-1} \circ \cdots \circ \phi_{t-m}(y), \qquad t \in T. \tag{2.17}$$

The random elements Y_t *are measurable with respect to the* σ*–field generated by* $\{\phi_{t-k} \mid k \geq 1\}$*. If* $(\tilde{Y}_t)_{t \in T}$ *is any other solution to* (2.14)*, then*

$$\mathrm{d}(\tilde{Y}_t, Y_t) \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty. \tag{2.18}$$

Moreover, in the case $T = \mathbb{Z}$ *the stationary solution to the* SRE (2.14) *is unique.*

Remarks 2.6.2. Note that there are in general many solutions to (2.14). Indeed, if $T = \mathbb{N}$, take any $z \in E$ in order to produce a solution $(\phi_{t-1}^{(t)}(z))_{t \in \mathbb{N}}$. The elements

$$\phi_{t-1}^{(t)}(z) = \phi_{t-1} \circ \cdots \circ \phi_0(z), \qquad t \geq 0,$$

are also called the forward iterates associated with the SRE (2.14), whereas the elements

$$\phi_{t-1}^{(m)}(z) = \phi_{t-1} \circ \cdots \circ \phi_{t-m}(z), \qquad m \geq 0,$$

with t fixed are called backward iterates. Thus relation (2.17) can be read as Y_t being the limit of its backward iterates, and relation (2.18) means that the forward iterates approach the trajectory of the unique stationary solution $(Y_t)_{t \in T}$ with an error decaying exponentially fast as $t \to \infty$. We also mention that the limit of the backward iterates is *irrespective* of $z \in E$, as follows from the limit relation (2.20) below. In general, the stationary distribution associated with the SRE (2.14) cannot be determined analytically.

The limit relation (2.18) provides a mean for the simulation of the stationary solution $(Y_t)_{t \in \mathbb{N}}$: take an arbitrary $z \in E$ and set $\tilde{Y}_{t+1} = \phi_t(\tilde{Y}_t)$ for $t \geq 0$. Then $\mathrm{d}(\tilde{Y}_t, Y_t) \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$.

If we have $T = \mathbb{N}$, there is more than one stationary solution to (2.14). One can however show that the law of the stationary solutions is unique; see Corollary 3.3 of Bougerol [20]. To illustrate this statement, consider a causal autoregressive process of order 1 with index set $T = \mathbb{N}$ and iid innovations $Z_t \sim \mathcal{N}(0, 1)$, i.e.,

$$X_{t+1} = \phi_t(X_t), \qquad t \in \mathbb{N}, \tag{2.19}$$

where $\phi_t(x) = \beta x + Z_{t+1}$ and $|\beta| < 1$. Then for *any* random variable $Y_0 \sim \mathcal{N}(0, (1-\beta^2)^{-1})$ which is independent of (Z_t), the sequence (Y_t) defined by

$$Y_t = \begin{cases} Y_0, & t = 0, \\ \beta^t Y_0 + \sum_{k=0}^{t-1} \beta^k Z_{t-k}, & t \geq 1, \end{cases}$$

provides a stationary solution of (2.19). □

Proof of Theorem 2.6.1. For the existence of a stationary ergodic solution it is enough to prove that the limit on the right–hand side of (2.17) exists; it is then obvious that $(Y_t)_{t\in\mathbb{N}}$ is a solution of the SRE $X_{t+1} = \phi_t(X_t)$. Indeed, the condition $\mathbb{E}[\log^+ \Lambda(\phi_t)] < \infty$ implies that $\Lambda(\phi_t) < \infty$ a.s., and hence ϕ_t is continuous a.s. From this we deduce

$$\phi_t(Y_t) = \phi_t\Big(\lim_{m\to\infty} \phi_{t-1}^{(m)}(y)\Big) = \lim_{m\to\infty} \phi_t^{(m+1)}(y) = Y_{t+1} \quad \text{a.s.}$$

The fact that (Y_t) is stationary ergodic and that Y_t is measurable with respect to the σ–field generated by $\{\phi_{t-k} \mid k \geq 1\}$ is a consequence of Corollary 2.1.3.

We now show that $\phi_{t-1}^{(m)}(y)$ converges as $m \to \infty$. Due to the stationarity of (ϕ_t), it is sufficient to show that $\phi_0^{(m)}(y)$ converges. By virtue of relation (2.13), for all $m_1, m_2 \in \mathbb{N}$,

$$\log \Lambda(\phi_0^{(m_1+m_2)}) = \log \Lambda(\phi_0^{(m_1)} \circ \phi_{-m_1}^{(m_2)}) \leq \log \Lambda(\phi_0^{(m_1)}) + \log \Lambda(\phi_{-m_1}^{(m_2)}).$$

Thus the sequence $(\log \Lambda(\phi_0^{(m)}))_{m\in\mathbb{N}}$ (together with the backshift operator on the space of doubly–infinite sequences of Lipschitz functions on E) is subadditive; see e.g. Section 10.7 in Dudley [43] for a detailed treatment of the notion of a subadditive sequence. Now, as a consequence of Kingman's subadditive ergodic theorem [74], with

$$\rho = \inf_{m\geq 1} \frac{1}{m} \mathbb{E}[\log \Lambda(\phi_0^{(m)})]$$

the following limit relations hold true:

$$\lim_{m\to\infty} \frac{1}{m} \mathbb{E}[\log \Lambda(\phi_0^{(m)})] = \lim_{m\to\infty} \frac{1}{m} \log \Lambda(\phi_0^{(m)}) = \rho \quad \text{a.s.} \tag{2.20}$$

Note that (2.20) is valid for *any* stationary ergodic sequence (ϕ_t) fulfilling $\mathbb{E}[\log^+ \Lambda(\phi_0)] < \infty$ and that condition (2.16) implies in particular $\rho < 0$. Also observe that $\mathbb{E}[\log^+ \mathrm{d}(\phi_0(y), y)] < \infty$ together with the Borel–Cantelli lemma implies

$$\frac{1}{m} \log^+ \mathrm{d}(\phi_{-m}(y), y) \xrightarrow{\text{a.s.}} 0, \qquad m \to \infty. \tag{2.21}$$

Now, accounting for (2.20) and (2.21), we have

$$\begin{aligned}
&\limsup_{m\to\infty} \frac{1}{m} \log \mathrm{d}(\phi_0^{(m+1)}(y), \phi_0^{(m)}(y)) \\
&= \limsup_{m\to\infty} \frac{1}{m} \log \mathrm{d}(\phi_0^{(m)}(\phi_{-m}(y)), \phi_0^{(m)}(y)) \\
&\le \limsup_{m\to\infty} \frac{1}{m} \log \Lambda(\phi_0^{(m)}) + \limsup_{m\to\infty} \frac{1}{m} \log \mathrm{d}(\phi_{-m}(y), y) \\
&= \rho \quad \text{a.s.}
\end{aligned}$$

Since $\rho < 0$, this shows that $(\phi_0^{(m)}(y))_{m\in\mathbb{N}}$ is a Cauchy sequence a.s., and thus converges due to the completeness of E. Concerning the limit relation (2.18), we first note that $t^{-1}\log \Lambda(\phi_{t-1}^{(t-1)}) \xrightarrow{\text{a.s.}} \rho$ by Kingman's subadditive ergodic theorem. This entails

$$\Lambda(\phi_{t-1}^{(t-1)}) \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty.$$

Using definition (2.12), we obtain

$$\mathrm{d}(Y_t, \tilde{Y}_t) = \mathrm{d}(\phi_{t-1}^{(t-1)}(Y_0), \phi_{t-1}^{(t-1)}(\tilde{Y}_0)) \le \Lambda(\phi_{t-1}^{(t-1)})\, \mathrm{d}(Y_0, \tilde{Y}_0) \xrightarrow{\text{e.a.s.}} 0.$$

It remains to demonstrate the uniqueness of the stationary solution (Y_t). If $(\check{Y}_t)$ is yet another stationary solution of the SRE (2.14), then by (2.12) and the triangle inequality,

$$\begin{aligned}
\mathrm{d}(\check{Y}_t, Y_t) &= \mathrm{d}(\phi_{t-1}^{(m)}(\check{Y}_{t-m}),\ \phi_{t-1}^{(m)}(Y_{t-m})) \le \Lambda(\phi_{t-1}^{(m)})\, \mathrm{d}(\check{Y}_{t-m}, Y_{t-m}) \\
&\le \Lambda(\phi_{t-1}^{(m)})(\mathrm{d}(\check{Y}_{t-m}, y) + \mathrm{d}(y, Y_{t-m})). \qquad (2.22)
\end{aligned}$$

The relation (2.20) with $\rho < 0$ and the stationarity of (ϕ_t) imply

$$\Lambda(\phi_{t-1}^{(m)}) \xrightarrow{\text{e.a.s.}} 0, \qquad m \to \infty.$$

Since $(\mathrm{d}(\check{Y}_{t-m}, y))_{m\in\mathbb{N}}$ and $(\mathrm{d}(y, Y_{t-m}))_{m\in\mathbb{N}}$ are stationary, a Slutsky argument shows that the right–hand side of (2.22) converges to zero in probability. For this reason, $\mathbb{P}(\mathrm{d}(\check{Y}_t, Y_t) = 0) = 1$, and $(\check{Y}_t)_{t\in T}$ and $(Y_t)_{t\in T}$ are indeed indistinguishable. □

Example 2.6.3. In order to illustrate Theorem 2.6.1, we apply it to autoregressive processes of order 1. Let (Z_t) be an iid sequence with $\mathbb{E}|Z_0| < \infty$, and consider the SRE

$$X_{t+1} = \beta X_t + Z_{t+1}, \qquad t \in Z, \qquad (2.23)$$

on $\mathbb{R}$. Thus $\phi_t(x) = \beta x + Z_{t+1}$ and $\Lambda(\phi_0^{(r)}) = |\beta|^r$ for all $r \ge 1$. Also note that $\mathbb{E}[\log^+ |\phi_0(0)|] = \mathbb{E}(\log^+ |Z_1|) < \infty$ due to $\mathbb{E}|Z_0| < \infty$. According to Theorem 2.6.1 the condition $|\beta| < 1$ is sufficient for the existence and uniqueness of a stationary solution (Y_t) to (2.23). From (2.17) we have the representation

$$Y_t = \lim_{m\to\infty} \phi_{t-1} \circ \cdots \circ \phi_{t-m}(0) = \lim_{m\to\infty} \sum_{k=0}^{m-1} \beta^k Z_{t-k} = \sum_{k=0}^{\infty} \beta^k Z_{t-k} \quad \text{a.s.}$$

The sequence (Y_t) is stationary ergodic and Y_t is measurable with respect to the σ–algebra generated by $\{Z_t, Z_{t-1}, \ldots,\}$ for each $t \in \mathbb{Z}$, or in other words, the solution (Y_t) is causal, i.e., Y_t is a function of the past and present innovations. This result is well–known, see e.g. or Brockwell and Davis [29]. However, Theorem 2.6.1 does not capture the fact that there are also unique stationary ergodic solutions when $|\beta| > 1$, because these solutions are *non–*causal. □

The following result relates the solutions of a SRE associated with a stationary ergodic sequence (ϕ_t) of Lipschitz maps with the solutions of a certain "perturbed" SRE.

Theorem 2.6.4. *Let $(B, \|\cdot\|)$ be a separable Banach space and (ϕ_t) be a stationary ergodic sequence of Lipschitz maps B into B. Impose:*

S.1 $\mathbb{E}(\log^+ \|\phi_0(0)\|) < \infty$.

S.2 $\mathbb{E}[\log^+ \Lambda(\phi_0)] < \infty$ *and for some integer* $r \geq 1$,

$$\mathbb{E}[\log \Lambda(\phi_0^{(r)})] = \mathbb{E}[\log \Lambda(\phi_0 \circ \cdots \circ \phi_{-r+1})] < 0.$$

Assume that $\mathbb{E}(\log^+ \|Y_0\|) < \infty$ *for the stationary solution* $(Y_t)_{t\in\mathbb{N}}$ *of the* SRE *associated with* (ϕ_t) *given by* (2.17). *Let* $(\hat{\phi}_t)_{t\in\mathbb{N}}$ *be a sequence of Lipschitz maps such that*

S.3 $\|\hat{\phi}_t(0) - \phi_t(0)\| \xrightarrow{\text{e.a.s.}} 0$ *and* $\Lambda(\hat{\phi}_t - \phi_t) \xrightarrow{\text{e.a.s.}} 0$ *as* $t \to \infty$.

Then for every solution $(\hat{Y}_t)_{t\in\mathbb{N}}$ *of the perturbed* SRE

$$X_{t+1} = \hat{\phi}_t(X_t), \qquad t \in \mathbb{N},$$

one has that

$$\|\hat{Y}_t - Y_t\| \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty. \tag{2.24}$$

Proof. Note that the sequence (ϕ_t) fulfills the conditions of Theorem 2.6.1 with d the metric induced by $\|\cdot\|$. It is sufficient to demonstrate $\|\hat{Y}_{s+kr} - Y_{s+kr}\| \xrightarrow{\text{e.a.s.}} 0$ as $k \to \infty$ for each $s \in [0, r)$; indeed, the latter limit relation implies $\|\hat{Y}_t - Y_t\| \leq \sum_{s=0}^{r-1} \|\hat{Y}_{s+r[t/r]} - Y_{s+r[t/r]}\| \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$. To begin with, we establish the auxiliary results

$$d_t := \|\hat{\phi}_t^{(r)}(0) - \phi_t^{(r)}(0)\| \xrightarrow{\text{e.a.s.}} 0 \quad \text{and} \quad e_t := \Lambda(\hat{\phi}_t^{(r)} - \phi_t^{(r)}) \xrightarrow{\text{e.a.s.}} 0, \tag{2.25}$$

as $t \to \infty$, i.e., the condition S.3 is fulfilled for the r–fold convolutions $\hat{\phi}_t^{(r)}$ and $\phi_t^{(r)}$. Note that the limit relation (2.25) is true if $r = 1$ by virtue of condition S.3. The proof of (2.25) goes by induction on r. Indeed, by the triangle inequality for any $m \geq 1$,

$$\|\hat\phi_t^{(m+1)}(0)-\phi_t^{(m+1)}(0)\| = \|\hat\phi_t\circ\hat\phi_{t-1}^{(m)}(0)-\phi_t\circ\phi_{t-1}^{(m)}(0)\|$$
$$\le \|\hat\phi_t\circ\hat\phi_{t-1}^{(m)}(0)-\hat\phi_t\circ\phi_{t-1}^{(m)}(0)\| + \|(\hat\phi_t-\phi_t)\circ\phi_{t-1}^{(m)}(0)-(\hat\phi_t-\phi_t)(0)\|$$
$$+\|(\hat\phi_t-\phi_t)(0)\|$$
$$\le \Lambda(\hat\phi_t)\|\hat\phi_{t-1}^{(m)}(0)-\phi_{t-1}^{(m)}(0)\| + \Lambda(\hat\phi_t-\phi_t)\|\phi_{t-1}^{(m)}(0)\| + \|\hat\phi_t(0)-\phi_t(0)\|.$$

The e.a.s. convergence to zero of the left–hand side is a consequence of $\Lambda(\hat\phi_t)\le\Lambda(\hat\phi_t-\phi_t)+\Lambda(\phi_t)$, $\mathbb{E}[\log^+\Lambda(\phi_0)]<\infty$, the limit relation $\|\hat\phi_{t-1}^{(\ell)}(0)-\phi_{t-1}^{(\ell)}(0)\|\xrightarrow{\text{e.a.s.}}0$ for every $\ell\in[0,m]$, and $\mathbb{E}(\log^+\|\phi_{t-1}^{(m)}(0)\|)<\infty$ together with repeated application of Proposition 2.5.1. Exploiting the submultiplicativity of Λ, one finds in a similar way that

$$\Lambda(\hat\phi_t^{(m+1)}-\phi_t^{(m+1)}) \le \Lambda(\hat\phi_t)\Lambda(\hat\phi_{t-1}^{(m)}-\phi_{t-1}^{(m)}) + \Lambda(\hat\phi_t-\phi_t)\Lambda(\phi_{t-1}^{(m)}) \xrightarrow{\text{e.a.s.}} 0,$$

and thus (2.25) is established.

If we set $\hat c_t=\Lambda(\hat\phi_t^{(r)})$ and take into account that every map ϕ on E fulfills $\|\phi(x)\|\le\|\phi(x)-\phi(0)\|+\|\phi(0)\|\le\Lambda(\phi)\|x\|+\|\phi(0)\|$, $x\in B$, we obtain

$$\|\hat Y_{s+kr}-Y_{s+kr}\| = \|\hat\phi_{s+kr-1}^{(r)}(\hat Y_{s+(k-1)r})-\phi_{s+kr-1}^{(r)}(Y_{s+(k-1)r})\|$$
$$\le \|\hat\phi_{s+kr-1}^{(r)}(\hat Y_{s+(k-1)r})-\hat\phi_{s+kr-1}^{(r)}(Y_{s+(k-1)r})\|$$
$$+\|\hat\phi_{s+kr-1}^{(r)}(Y_{s+(k-1)r})-\phi_{s+kr-1}^{(r)}(Y_{s+(k-1)r})\|$$
$$\le \Lambda(\hat\phi_{s+kr-1}^{(r)})\,\|\hat Y_{s+(k-1)r}-Y_{s+(k-1)r}\| + \Lambda(\hat\phi_{s+kr-1}^{(r)}-\phi_{s+kr-1}^{(r)})\,\|Y_{s+(k-1)r}\|$$
$$+\|\hat\phi_{s+kr-1}^{(r)}(0)-\phi_{s+kr-1}^{(r)}(0)\|$$
$$= \hat c_{s+kr-1}\,\|\hat Y_{s+(k-1)r}-Y_{s+(k-1)r}\| + (e_{s+kr-1}\,\|Y_{s+(k-1)r}\|+d_{s+kr-1}).$$

Iterating the latter inequality until $k=1$, we receive the final estimate

$$\|\hat Y_{s+kr}-Y_{s+kr}\| \le \left(\prod_{\ell=1}^{k}\hat c_{s+\ell r-1}\right)\|\hat Y_s-Y_s\| + \sum_{\ell=1}^{k}\left(\prod_{i=\ell+1}^{k}\hat c_{s+ir-1}\right)(e_{s+\ell r-1}\,\|Y_{s+(\ell-1)r}\|+d_{s+\ell r-1}). \tag{2.26}$$

We show that the right–hand side of the latter inequality tends to zero e.a.s. Set $c_t=\Lambda(\phi_t^{(r)})$. An application of the monotone convergence theorem to $\mathbb{E}[\log^-(c_0+\epsilon)]$ and an application of the dominated convergence theorem to $\mathbb{E}[\log^+(c_0+\epsilon)]$ show $\mathbb{E}[\log(c_0+\epsilon)]\to\mathbb{E}[\log c_0]$ as $\epsilon\downarrow 0$. Thus there is $\epsilon_0 > 0$ such that $\mathbb{E}[\log(c_0 + \epsilon_0)] < 0$, and from Lemma 2.5.5 it follows $\prod_{\ell=1}^{k}(c_{s+\ell r-1} + \epsilon_0) \xrightarrow{\text{e.a.s.}} 0$ as $k \to \infty$. This limit relation together with the facts that $\hat{c}_t = \Lambda((\hat{\phi}_t^{(r)} - \phi_t^{(r)}) + \phi_t^{(r)}) \le e_t + c_t$ and a.s. $e_t \le \epsilon_0$ for all but finitely many t's leads to

$$\prod_{\ell=1}^{k} \hat{c}_{s+\ell r-1} \le \prod_{\ell=1}^{k}(e_{s+\ell r-1} + c_{s+\ell r-1}) \xrightarrow{\text{e.a.s.}} 0, \qquad k \to \infty, \tag{2.27}$$

and shows that $(\prod_{\ell=1}^{k} \hat{c}_{s+\ell r-1})\|\hat{Y}_s - Y_s\| \xrightarrow{\text{e.a.s.}} 0$. As regards the second term in (2.26), we use Proposition 2.5.1 together with (2.25) and $\mathbb{E}(\log^+ \|Y_0\|) < \infty$ to prove that $p_{s+\ell r-1} := e_{s+\ell r-1}\, \|Y_{s+(\ell-1)r}\| + d_{s+\ell r-1} \xrightarrow{\text{e.a.s.}} 0$ when $\ell \to \infty$. This limit result and (2.27) imply the existence of a random variable a_0 and a number $0 < \gamma < 1$ so that $p_{s+\ell r-1} \le a_0 \gamma^\ell$ for all $\ell \ge 1$. From this bound we obtain

$$\sum_{\ell=1}^{k}\left(\prod_{i=\ell+1}^{k} \hat{c}_{s+ir-1}\right) p_{s+\ell r-1} \le a_0 \sum_{\ell=1}^{k}\left(\prod_{i=\ell+1}^{k} \hat{c}_{s+ir-1}\right)\gamma^\ell.$$

Since the map $\delta \mapsto \mathbb{E}[\log(c_0 + \delta)]$ is continuous and increasing and $\mathbb{E}(\log c_0) < 0$, there exists a $\delta_0 > 0$ such that $0 > \mathbb{E}[\log(c_0 + \delta_0)] > \log\gamma$. Set $\tilde{\hat{c}}_t = \hat{c}_t + \delta_0$ and $\tilde{c}_t = c_t + \delta_0$, and note that we are done if we can show $\sum_{\ell=1}^{k}(\prod_{i=\ell+1}^{k} \tilde{\hat{c}}_{s+ir-1})\gamma^\ell \xrightarrow{\text{e.a.s.}} 0$. Since $(\tilde{c}_t)$ is stationary ergodic and $\mathbb{E}(\log \tilde{c}_0) > \log\gamma$, we have that $\gamma^\ell(\prod_{i=1}^{\ell} \tilde{c}_{s+ir-1})^{-1} \xrightarrow{\text{e.a.s.}} 0$. Using a Taylor argument together with $\tilde{\hat{c}}_t, \tilde{c}_t \ge \delta_0$, we conclude

$$|\,(\tilde{\hat{c}}_{s+ir-1})^{-1} - (\tilde{c}_{s+ir-1})^{-1}\,| \le (1/\delta_0^2)\,|\tilde{\hat{c}}_{s+ir-1} - \tilde{c}_{s+ir-1}| \xrightarrow{\text{e.a.s.}} 0, \qquad i \to \infty,$$

and therefore the identical arguments as used for (2.27) yield that also $\gamma^\ell(\prod_{i=1}^{\ell} \tilde{\hat{c}}_{s+ir-1})^{-1} \xrightarrow{\text{e.a.s.}} 0$. This implies the existence of a random variable b_0 with $\gamma^\ell(\prod_{i=1}^{\ell} \tilde{\hat{c}}_{s+ir-1})^{-1} \le b_0$ for all $\ell \ge 1$, so that

$$\sum_{\ell=1}^{k}\left(\prod_{i=\ell+1}^{k} \tilde{\hat{c}}_{s+ir-1}\right)\gamma^\ell \le b_0\, k \prod_{i=1}^{k} \tilde{\hat{c}}_{s+ir-1} \xrightarrow{\text{e.a.s.}} 0, \qquad \ell \to \infty.$$

For the last step we used that $\prod_{i=1}^{k} \tilde{\hat{c}}_{s+ir-1} \xrightarrow{\text{e.a.s.}} 0$, which is a consequence of $\mathbb{E}(\log \tilde{c}_0) < 0$ and $|\tilde{\hat{c}}_t - \tilde{c}_t| \xrightarrow{\text{e.a.s.}} 0$ together with the same arguments as applied in the derivation of (2.27). This completes the proof. □

Example 2.6.5 (Continuation of Example 2.6.3). Consider the autoregressive equation

$$X_{t+1} = \beta_t X_t + Z_{t+1}, \qquad t \in \mathbb{N}, \tag{2.28}$$

with time dependent coefficients $\beta_t \xrightarrow{\text{e.a.s.}} \beta \in (-1, 1)$ as $t \to \infty$ and (Z_t) an iid sequence with $\mathbb{E}|Z_0| < \infty$. Assume that $(\hat{Y}_t)_{t\in\mathbb{N}}$ is a certain solution of

(2.28). Then by incorporating the results of Example 2.6.3 and by an application of Theorem 2.6.4, we immediately receive that

$$\left|\hat{Y}_t - \sum_{k=0}^{\infty} \beta^k Z_{t-k}\right| \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty.$$

We mention that the condition $\mathbb{E}(\log^+ |\sum_{k=0}^{\infty} \beta^k Z_{t-k}|) < \infty$ needed for the applicability of Theorem 2.6.4 is met by virtue of $\mathbb{E}(\sum_{k=0}^{\infty} |\beta|^k |Z_{t-k}|) = (1 - |\beta|)^{-1} \mathbb{E}|Z_0| < \infty$. □

3

Financial Time Series: Facts and Models

Financial time series analysis deals with the analysis of data collected on financial markets. Its main goal is to gain reliable information in order to make rational decisions about the future. In the language of statistics, one seeks to predict the future behavior of financial markets. Although the economic sciences have gained insight into various phenomena related to economic activity, it is widely accepted that it is difficult, if not impossible, to precisely predict future economic indicators or prices of financial assets. This scepticism is based on the argument that these quantities are determined by a variety of interdependent factors, whose evolution and influence are difficult to quantify and to measure, and some of which might even be unknown. For this reason, stochastic models are often considered as more realistic than deterministic models. Whatever preferences one has, one should always keep in mind that a model for the description of financial markets hardly ever reflects total reality. Any model can only describe a few aspects of financial or economic processes.

Although every model is likely to be incorrect, one would like to distinguish between "better" and "worse" models. A statistical model is commonly regarded as a good approximation to reality if it yields a "satisfactory" fit to observed data. Raw financial data consists of a time series of prices P_t, $t = 0, \ldots, n$, of a certain asset. This financial asset could e.g. be the stock of a certain company or a stock index, a foreign currency, or a commodity, such as gold or oil. We assume that the times of observations are equidistant, so that we can avoid the particular difficulties of high–frequency data, where observations are usually irregularly spaced over time, see e.g. the book by Dacorogna et al. [33].

It is a common technique (Taylor [126]) to take log–differences so that the observations are transformed into so–called "log–returns"

$$X_t = \log P_t - \log P_{t-1} = \log \left(1 + \frac{P_t - P_{t-1}}{P_{t-1}} \right).$$

A Taylor series argument shows that log–returns are almost indistinguishable from relative returns, provided the latter are small, i.e.,

$$X_t \approx \frac{P_t - P_{t-1}}{P_{t-1}}\,.$$

Log–returns are often preferred to relative returns because they are additive with respect to aggregation over different time periods, a property which relative returns do not share.

One of the most fundamental working hypotheses in financial time series analysis is the (strict) stationarity of log–return data, at least over "appropriate" (not "too long") time periods. See Section 2.1 for a mathematical definition of the stationarity of a stochastic process. Only recently the hypothesis of stationarity has been questioned. Since this monograph is focused rather on theoretical questions than on data analysis, we refer the reader to Mikosch and Stărică [101, 100] or Mikosch [97] for a detailed discussion of the non–stationarity issue. In this monograph we work under the assumption of stationarity.

In what follows, we present some of the "stylized facts" of financial log–return time series data. As an illustration, we include a chart with the daily closing prices of the New York Stock Exchange Composite Index (NYSE Composite) in Figure 3.1. The corresponding log–returns are provided by Figure 3.2. The NYSE Composite measures the performance of all common stocks listed on the NYSE. We mention that access to information about the composition of this index and to historical data is granted under `http://www.nyse.com/marketinfo/marketinfo.html`.

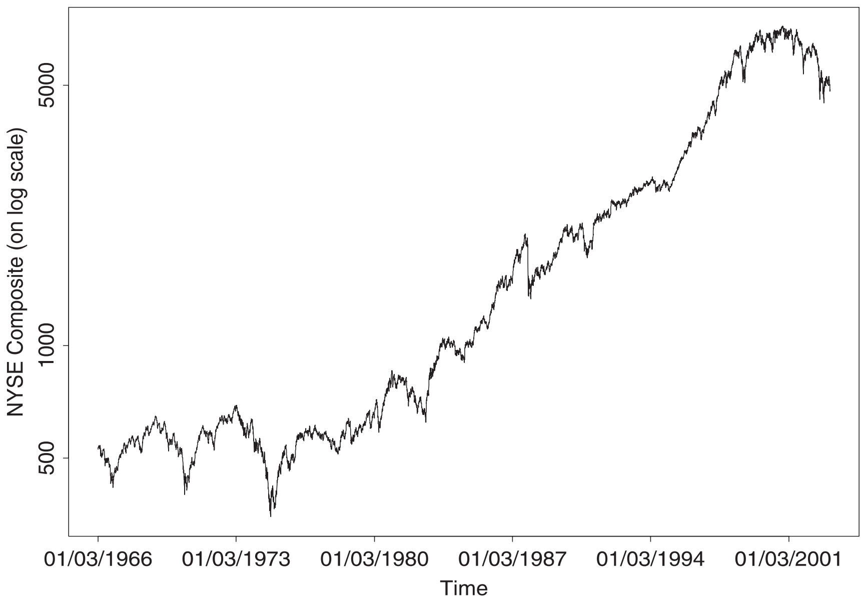

Fig. 3.1. NYSE Composite closing prices from January 3, 1966 – January 28, 2003.

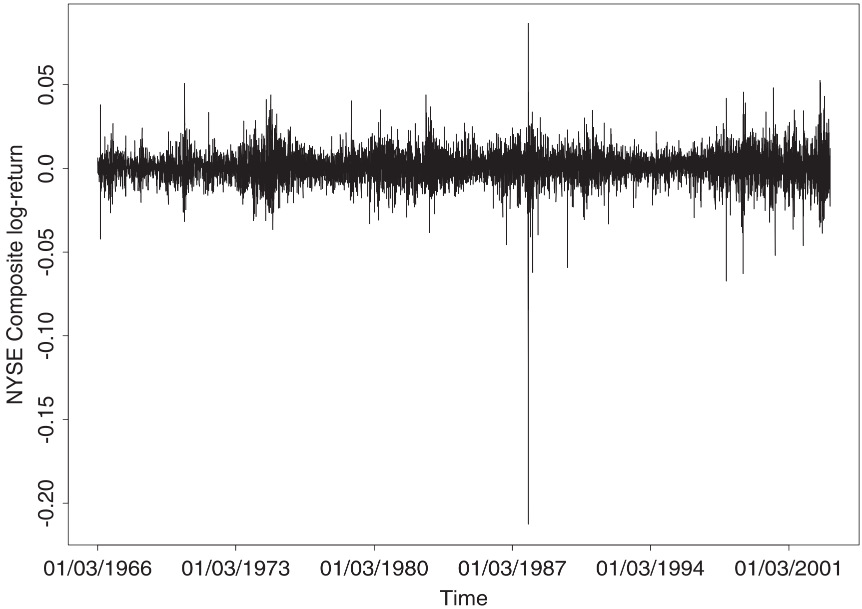

Fig. 3.2. NYSE Composite log–returns from January 3, 1966 – January 28, 2003.

3.1 Stylized Facts of Financial Log–return Data

3.1.1 Uncorrelated Observations

In classical time series analysis, the dependence between consecutive observations is most often measured via the autocovariances or autocorrelations, see e.g. the books by Brockwell and Davis [28, 29] or Fuller [52], which we consider as comprehensive references. Let (X_t) be a stationary stochastic process with finite variance. Then the autocovariance $\gamma_X(h)$ and autocorrelation $\rho_X(h)$ at lag $h \in \mathbb{Z}$ is the covariance and correlation, respectively, between observations which are h time steps apart, i.e.,

$$\gamma_X(h) = \mathrm{Cov}(X_0, X_h) \quad \text{and} \quad \rho_X(h) = \frac{\mathrm{Cov}(X_0, X_h)}{\mathrm{Var}(X_0)} .$$

The functions $\gamma_X(\cdot)$ and $\rho_X(\cdot)$ defined on the integers are called autocovariance function (ACVF) and autocorrelation function (ACF), respectively. The sample counterparts of the autocovariances and autocorrelation, the sample autocovariances and sample autocorrelations, are the estimators

$$\gamma_{n,X}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_t - \overline{X})(X_{t+|h|} - \overline{X}) \quad \text{and} \quad \rho_{n,X}(h) = \frac{\gamma_{n,X}(h)}{\gamma_{n,X}(0)} ,$$

where $\overline{X} = n^{-1} \sum_{t=1}^{n} X_t$ is the sample mean of the data $(X_1, \ldots, X_n)$. When $\gamma_{n,X}(\cdot)$ or $\rho_{n,X}(\cdot)$ are regarded as functions, one speaks of the sample autocovariance function (SACVF) and sample autocorrelation function (SACF). Provided (X_t) is also ergodic, a straightforward application of the ergodic theorem yields for every fixed $h \in \mathbb{Z}$ that

$$\gamma_{n,X}(h) \xrightarrow{\text{a.s.}} \gamma_X(h) \quad \text{and} \quad \rho_{n,X}(h) \xrightarrow{\text{a.s.}} \rho_X(h), \qquad n \to \infty.$$

This means that the estimators $\gamma_{n,X}(h)$ and $\rho_{n,X}(h)$ are strongly consistent for $\gamma_X(h)$ and $\rho_X(h)$. Real–life log–return data often exhibit the following peculiar behavior:

- $\rho_{n,X}(h)$ vanishes at all lags $h \neq 0$, except perhaps at the first lag $|h| = 1$.
- In contrast, the sample autocorrelations $\rho_{n,|X|}(h)$ of the absolute log–returns are different from zero for a large number of lags h.

The sample autocorrelations belonging to the NYSE Composite data exactly show this behavior, see Figure 3.3.

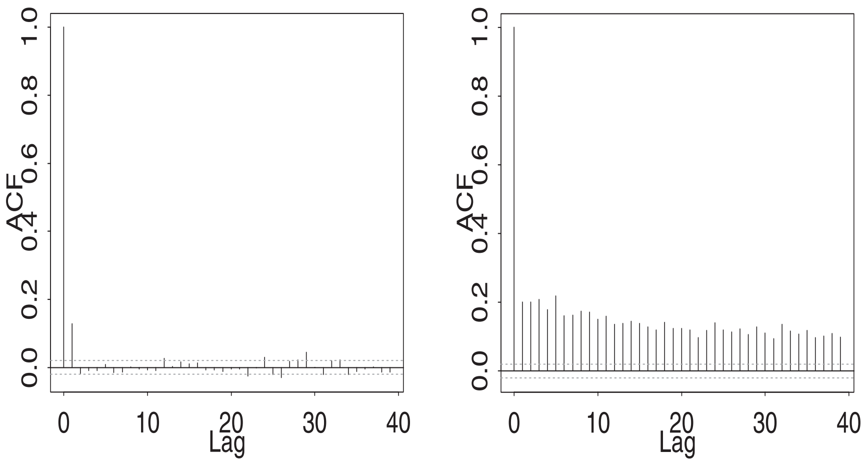

Fig. 3.3. Sample autocorrelations for NYSE Composite log–returns and absolute log–returns. At every *fixed* lag the dotted horizontal lines define an asymptotic 95% acceptance region of the null hypothesis of iid finite variance noise. For a derivation of this region, we refer to Section 7.2 in Brockwell and Davis [29].

Therefore, a reasonable stochastic model for log–returns seems to be a stationary white noise process (U_t), which is *not* iid. We recall that a process (U_t) is a white noise process (or sequence) if (U_t) has constant mean and constant finite variance, and autocovariances $\gamma_U(h) = 0$ at lags $h \neq 0$.

3.1.2 Time–varying Volatility (Conditional Heteroscedasticity)

As a matter of fact, financial markets react "nervously" in the presence of political disorders, economic crises, war or fear of war, or in the event of a major natural catastrophe or man–made disaster, which is believed to threaten human society. During such stress periods the prices of financial assets usually fluctuate strongly. In statistical terms, the conditional variance given the past, i.e., $\mathrm{Var}[X_t \,|\, X_{t-1}, X_{t-2}, \ldots]$ is not constant over time and the underlying stochastic process (X_t) is conditionally heteroscedastic. Econometricians would also say that the volatility

$$\sigma_t = (\mathrm{Var}[X_t \,|\, X_{t-1}, X_{t-2}, \ldots])^{1/2}$$

changes over time. The human eye easily detects stress periods in the NYSE Composite time series of Figure 3.2. The corresponding historical events, which are presumed to have triggered these market conditions, at least to some extent, are as follows: the Prague spring invasion (summer 1968), the oil crisis of 1974, the fall of the Shah of Persia (spring 1980), the stock market crash in fall 1987, the second Gulf War (summer 1990), the economic crisis in Southeast Asia (fall 1997), the Russian financial crisis (fall 1998), and the September WTC attacks (2001). We are aware of the fact that our list merely cites the most incisive events and is far from being complete.

Without any model assumption it is rather difficult to statistically estimate the unobservable quantity σ_t, and any chosen nonparametric method can probably be put into question by a critically minded person. We decided to employ the exponentially weighted moving average (EWMA) as an estimator, i.e., we compute the squared volatility recursively by

$$\hat{\sigma}_t^2 = (1 - \lambda) X_t^2 + \lambda \hat{\sigma}_{t-1}^2$$

and starting value $\hat{\sigma}_0^2 = 0$, say. The exponent $0 < \lambda < 1$ can be regarded as a smoothing parameter, for it was shown by Gijbels et al. [54] that the EWMA can be interpreted as the Nadaraya–Watson kernel estimator of $\mathbb{E}[X_t^2 \,| \,|X_{t-1}, \ldots, X_0]$, with a kernel function that is zero in its positive argument. EWMA is also applied by RiskMetrics© [119]. The estimated volatility is graphed in Figure 3.4. For more sophisticated ways of nonparametric volatility estimation we refer to Bühlmann and McNeil [30].

3.1.3 Heavy–tailed and Asymmetric Unconditional Distribution

Many financial log–return time series seem to indicate that the unconditional distribution of X_t is heavy–tailed, i.e., not all moments of X_t exist. A QQ–plot of the negative and positive values of X_t against standard exponential quantiles in Figure 3.5 reveals that the left and right tails of NYSE Composite log–return data decay slightly slower than the tail of an exponential distribution. In order to investigate the tail of $|X_t|$, we fit a (generalized) Pareto tail

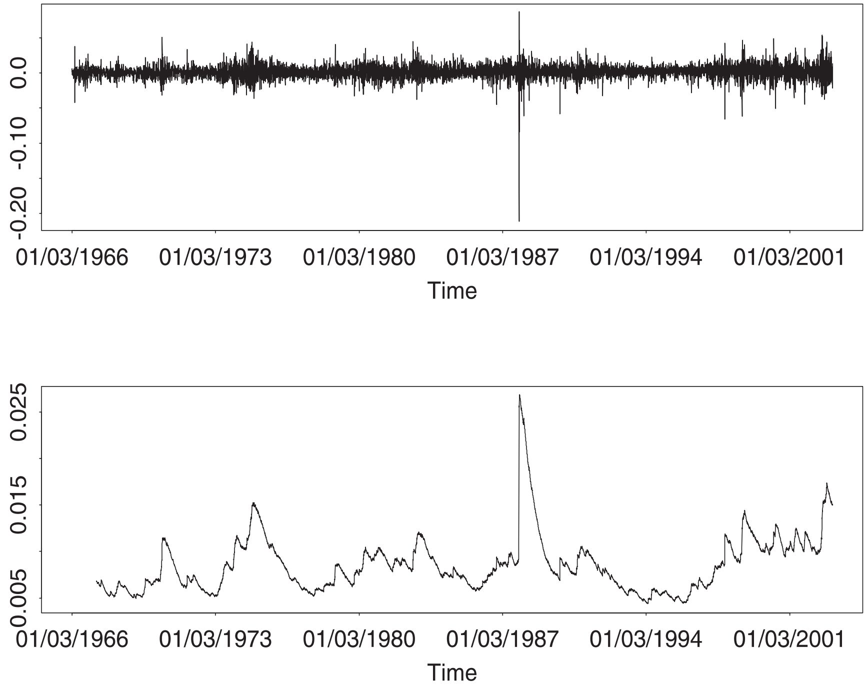

Fig. 3.4. NYSE Composite log–returns (top) and estimated volatility (bottom). For EWMA, the smoothing parameter $\lambda = 0.99$ was used. The first 300 values of $\hat{\sigma}_t$ have been discarded.

to the tail of the empirical distribution function of absolute NYSE Composite log–returns. Figure 3.6 indicates a nice fit. The value $\alpha = 3.31$ can be seen as an estimate for the tail index of the data generating model and seems to indicate that the tail of the distribution function of $|X_t|$, i.e., $\bar{F}(x) = \mathbb{P}(|X_t| > x)$, decays like a power function:

$$\bar{F}(x) \sim cx^{-\alpha}, \qquad x \to \infty.$$

Here the symbol $\sim$ is defined as follows: two nonnegative functions $f(x)$ and $g(x)$ on $\mathbb{R}_+$ (or $\mathbb{N}$) fulfill $f(x) \sim g(x)$ as $x \to \infty$ if

$$\frac{f(x)}{g(x)} \to 1, \qquad x \to \infty.$$

For other possible methods of tail index estimation we refer to Chapter 6 of Embrechts et al. [45]. The lower graph in Figure 3.5 indicates that the distribution of X_t might be asymmetric, i.e., the distributions of X_t and $-X_t$ differ in the tails. However, since we totally neglected the dependence, our ad hoc exploratory data analysis has to be interpreted with some care.

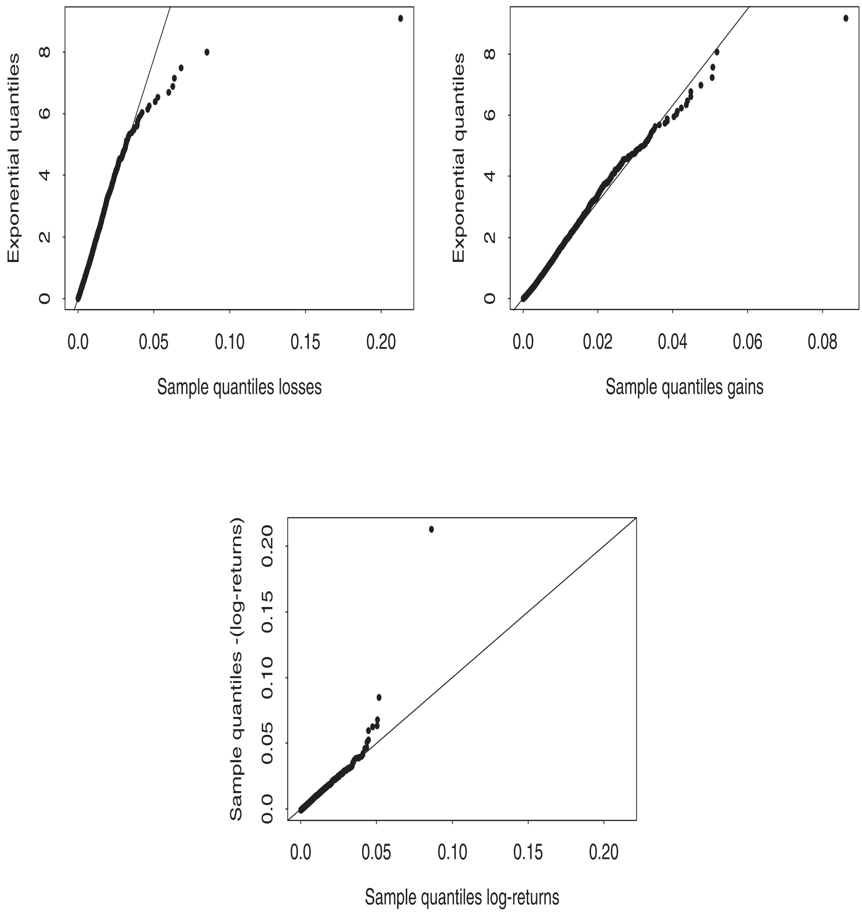

Fig. 3.5. *Top*: QQ–plots of negative (left) and positive (right) NYSE Composite log–return data against standard exponential. *Bottom*: Comparison of the distributions of X_t and of $-X_t$ via a QQ–plot. Due to symmetry of the QQ–plot it is sufficient to display the part with nonnegative first coordinate only.

3.1.4 Leverage Effects

There are theoretical considerations which suggest that volatility tends to respond asymmetrically to positive and negative log–returns. It is believed that this asymmetric behavior is caused by so–called leverage effects; see Nelson [106] for an accessible account on this subject. Outside the framework of a model there is no unique procedure for a statistical verification of the presence of leverage effects. As a very simple ad hoc method we propose to compare the conditional distribution of $|X_t|$ given $X_{t-1} > 0$ to the conditional distribution of $|X_t|$ given $X_{t-1} < 0$, which can be achieved by comparison of

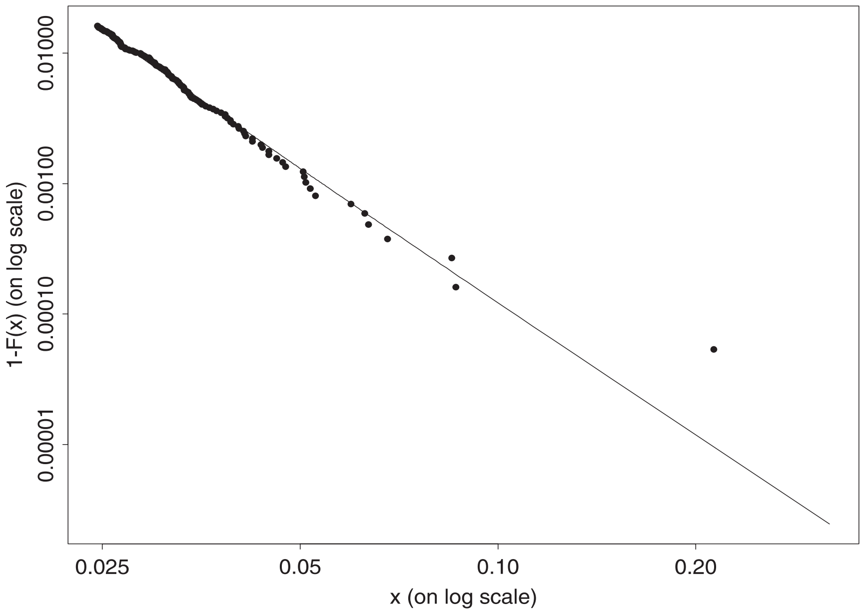

Fig. 3.6. Tail of the empirical distribution function of absolute NYSE Composite log–returns, evaluated at the 150 largest values. The solid line is a plot of the tail of the GPD (Generalized Pareto Distribution), $\bar{F}(x) = (1 + (\alpha\beta)^{-1}(x - \mu))^{-\alpha}$, with parameters $\alpha = 3.31$, $\mu = 0.00861$, $\beta = 0.00194$. The fit was obtained by application of the Splus function `gpd` of the EVIS Software package by McNeil [94].

the corresponding empirical quantiles. The resulting plot in Figure 3.7 should again be interpreted restrainedly by reason of unknown effects caused by the dependence between the sample quantiles of the two different conditional distributions. The QQ–plot does not contradict the economic hypothesis that a negative return is followed by volatilities which are larger than the volatilities following a positive return of the same absolute size.

3.2 ARMA Models

We introduce autoregressive moving average (ARMA) processes because they form the backbone of classical time series analysis and because squared GARCH processes can be interpreted as ARMA processes, see Section 4.2.2. ARMA processes are however not very suitable models for log–return time series, as we will see below. Recall that a stationary stochastic process (X_t) is said to be an ARMA(p, q) process (with mean zero) if it satisfies the difference equation

$$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + \vartheta_1 Z_{t-1} + \cdots + \vartheta_q Z_{t-q} + Z_t, \quad t \in \mathbb{Z}, \quad (3.1)$$

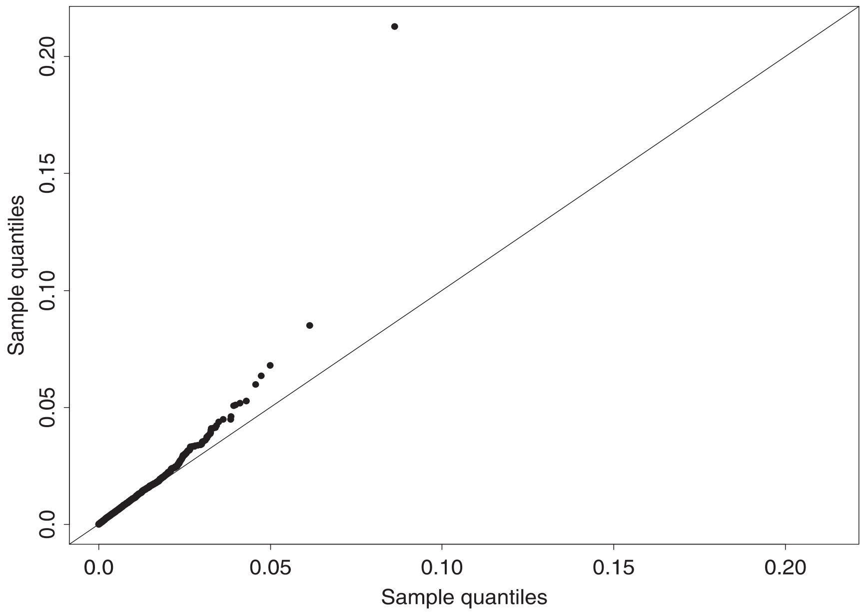

Fig. 3.7. Sample quantiles of the conditional distribution of $|X_t|$ given $X_{t-1} > 0$ (on x–axis) against the sample quantiles of the conditional distribution of $|X_t|$ given $X_{t-1} < 0$.

where $\varphi_1, \ldots, \varphi_p$, $\vartheta_1, \ldots \vartheta_q$ are real coefficients and where (Z_t) is a zero–mean iid sequence. Often the Z_t's are called innovations or noise. A stochastic process (X_t) is ARMA(p, q) with mean $\mu \in \mathbb{R}$ if $(X_t - \mu)$ is a zero–mean ARMA(p, q) process. Equivalently, one writes

$$\varphi(B)(X_t - \mu) = \vartheta(B) Z_t, \tag{3.2}$$

where B denotes the backshift operator and

$$\varphi(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p, \qquad \vartheta(z) = 1 + \vartheta_1 z + \cdots + \vartheta_q z^q$$

are the so–called characteristic polynomials. In the following we may assume without loss of generality that $\mu = 0$ and that the two polynomials have no common zeros since otherwise one can cancel out the corresponding factors of $\varphi(z)$ and $\vartheta(z)$ (see also Remark 1 in Chapter 3 of Brockwell and Davis [29]). It is well–known that the equation (3.1) admits a unique stationary ergodic solution (X_t) provided $\phi(z) \neq 0$ on $\{|z| = 1\}$ (Theorem 3.1.3 in Brockwell and Davis [29]).

In what follows, we restrict ourselves to causal ARMA processes, i.e., we assume that X_t can be written in the form

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \qquad t \in \mathbb{Z}, \tag{3.3}$$

where (ψ_j) is an absolutely summable real sequence with $\psi_0 = 1$. The causality assumption (3.3) of course entails stationarity; compare with Example 2.1.4. As a matter of fact, causality is equivalent to the property that the polynomial $\varphi(z)$ has no root in the unit disc, i.e., $\varphi(z) \neq 0$ for $|z| \leq 1$. In that case, the coefficients (ψ_j) of (3.3) are determined by the Taylor series representation of $\vartheta(z)/\varphi(z)$, i.e.,

$$\sum_{j=0}^{\infty} \psi_j z^j = \frac{\vartheta(z)}{\varphi(z)}, \qquad |z| \leq 1.$$

Another important property is invertibility. An ARMA(p,q) process is said to be invertible if there exists a sequence $(\pi_j)_{j\in\mathbb{N}}$ of absolutely summable constants such that

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \qquad t \in \mathbb{Z}.$$

In other words, the unobserved noise Z_t can be perfectly reconstructed from the present and past observations X_j, $j \leq t$. A causal ARMA(p,q) process is invertible if and only if the corresponding characteristic polynomials are such that the polynomial $\vartheta(z)$ has no root in the unit disc, and in that case $(\pi_j)_{j\in\mathbb{N}}$ is determined by the Taylor series representation of $\varphi(z)/\vartheta(z)$, i.e.,

$$\sum_{j=0}^{\infty} \pi_j z^j = \frac{\varphi(z)}{\vartheta(z)}, \qquad |z| \leq 1. \tag{3.4}$$

Causality and invertibility are usual assumptions in the classical treatment of the estimation of ARMA parameters, see Chapter 8 of Brockwell and Davis [29]. We mention that general models of the form (3.3) with an absolutely summable sequence (ψ_j) are also called causal linear processes (or MA(∞)). Therefore, causal ARMA processes form a subclass of linear processes.

ARMA models are important for several reasons. First of all, they are mathematically tractable. The determination of the parameter regions of stationarity, causality and invertibility are relatively simple and many relevant stochastic quantities, such as autocovariances or the spectral density, can be computed explicitly. Parameter estimation is well understood. Analogously to linear differential equations, the study of ARMA processes has led to the development of a beautiful theory, which nowadays can be regarded as fairly complete.

Secondly, the practical relevance of causal ARMA processes lies in the fact that they approximate general stationary processes in a certain sense. Indeed, to any linear process (3.3) one can find a causal ARMA(p,q) process with an error of approximation as small as we desire. This follows from the spectral representation theorem together with the fact that the set of spectral densities of causal ARMA processes is dense in the space of spectral densities of linear processes (with respect to the sup–norm). We refer to Chapter 4 of Brockwell and Davis [29] for the spectral theory of linear time series or to the books by Priestley [114] or Fuller [52]. See also Chapter 4 of this monograph for a brief summary of spectral theory. According to the Cramér–Wold decomposition, a general (not necessarily linear) stationary ergodic process (X_t) with mean zero and finite variance has unique representation

$$X_t = \sum_{j=0}^{\infty} \xi_j U_{t-j}, \qquad t \in \mathbb{Z}, \tag{3.5}$$

where $(\xi_j)_{j \geq 0}$ is a sequence of numbers with $\xi_0 = 1$, $\sum_{j=0}^{\infty} \xi_j^2 < \infty$ and (U_t) is a white noise sequence; consult Section 5.7 in Brockwell and Davis [29] for a proof of the Cramér–Wold decomposition and note that (U_t) will not be iid in general. From (3.5) we infer that there exists a linear process (Y_t) with identical second–order moment structure, i.e., with the same autocovariance function as (X_t), and even with Gaussian innovations. Hence causal ARMA models are at least suitable for capturing the second–order moment structure of an arbitrary stationary process. It was however shown by Bickel and Bühlmann [9] that causal ARMA models fail in general when it comes to approach the entire distribution of nonlinear processes. On the other hand, these two authors obtained the remarkable result that it is nevertheless impossible to sharply distinguish a nonlinear process from a high enough order MA process. Linear processes can exhibit "nonlinear behavior".

Many authors also admit white noise sequences (Z_t) in the definition (3.1) of an ARMA(p, q) model, but then not every ARMA process will be linear in the sense of definition (3.3). In this monograph we do not use such a general notion of ARMA models because it is only useful if one does not want to go beyond studying the second–order moment properties. Without any additional assumptions it is impossible to do statistical inference. This statement is nicely illustrated in Chapter 8. There it is shown that the so–called Whittle estimator applied to the squares of a GARCH$(1, 1)$ process, which obeys an ARMA$(1, 1)$ equation with a *non*–iid white noise sequence, has totally different statistical properties than the Whittle estimator of an ARMA$(1, 1)$ process with the same parameters and iid noise.

It is not difficult to give arguments against ARMA models for log–return data. A main reason why ARMA is not suitable for the description of log–returns lies in the fact that they do not allow conditional heteroscedasticity. Indeed, in a causal ARMA process the volatility is constant over time because

$$\mathrm{Var}(X_t \mid Z_{t-1}, Z_{t-2}, \ldots) = \mathrm{Var}(Z_0).$$

Furthermore, the only invertible causal ARMA(p, q) process with uncorrelated observations is $(X_t) = (Z_t)$, i.e., iid white noise. Recall that we indicated in Section 3.1.1 that financial log–returns do not seem to be iid at all. Hence we have to find models which are more appropriate than ARMA.

3.3 Conditionally Heteroscedastic Time Series Models

Many conditionally heteroscedastic time series models can be written in the form

$$X_t = \mu_t + \sigma_t Z_t, \qquad t \in \mathbb{Z}, \tag{3.6}$$

where (Z_t) is an iid sequence with $\mathbb{E}Z_0 = 0$ and $\mathrm{Var}(Z_0) = \mathbb{E}Z_0^2 = 1$, and (μ_t) and (σ_t) are stochastic processes which are assumed to depend only on the past, i.e., for every fixed $t \in \mathbb{Z}$ the random variables μ_{t+1}, σ_{t+1} are measurable with respect to the σ–field

$$\mathcal{F}_t = \sigma(Z_j\,, j \le t). \tag{3.7}$$

Hence X_t is $\mathcal{F}_t$–measurable for every t, which can be seen as an analogue to causality in ARMA models. Note that $\mu_t = \mathbb{E}(X_t|\mathcal{F}_{t-1})$ is the conditional mean of X_t given the past and that $|\sigma_t|$ can be interpreted as the volatility at time t since $\sigma_t^2 = \mathrm{Var}(X_t|\mathcal{F}_{t-1})$ by $\mathbb{E}Z_t^2 = 1$. In most models, $\sigma_t \ge 0$ a.s. Conditional heteroscedasticity emerges if (σ_t^2) is a nontrivial process. Observe that any causal ARMA process satisfies (3.6) with $\sigma_t = 1$. In financial applications the processes (μ_t) and (σ_t) are not directly observable. The autocovariances in model (3.6) can be obtained through the equations

$$\begin{aligned}\gamma_X(h) &= \mathrm{Cov}(X_0, X_{|h|}) \\ &= \mathbb{E}\left[\mathrm{Cov}(X_0, X_{|h|} \,|\, \mathcal{F}_0)\right] + \mathrm{Cov}\left[\mathbb{E}(X_0 \,|\, \mathcal{F}_0), \mathbb{E}(X_{|h|} \,|\, \mathcal{F}_0)\right] \\ &= \mathrm{Cov}\left[X_0, \mathbb{E}(X_{|h|} \,|\, \mathcal{F}_0)\right],\end{aligned}$$

which are valid if (X_t) is stationary and $\mathrm{Var}(X_0) < \infty$. In the special case of $\mu_t = \mu = \text{const.}$, we have $\mathbb{E}(X_{|h|} \,|\, \mathcal{F}_0) = \mathbb{E}[\,\mathbb{E}(X_{|h|} \,|\, \mathcal{F}_{|h|-1}) \,|\, \mathcal{F}_0\,] = \mu$ a.s. since $\mathcal{F}_0 \subset \mathcal{F}_{|h|-1}$ and $\mathbb{E}(X_{|h|} \,|\, \mathcal{F}_{|h|-1}) = \mu$ a.s. Hence, if $\mu_t = \mu$, then

$$\gamma_X(h) = 0, \qquad h \neq 0,$$

which means that (X_t) is a white noise process. We now provide some nontrivial and important examples.

3.3.1 AGARCH Models

Here we introduce the class of so–called asymmetric GARCH (AGARCH) models, which also contains ordinary GARCH. The aim of the present section is to collect the basic facts of AGARCH and its subclasses GARCH and ARCH.

A time series (X_t) is called an AGARCH(p, q) (asymmetric generalized autoregressive conditionally heteroscedastic) process if it satisfies

$$X_t = \sigma_t Z_t, \qquad t \in \mathbb{Z}, \tag{3.8}$$

where (Z_t) is a sequence of iid random variables with $\mathbb{E}Z_0 = 0$ and $\mathbb{E}Z_0^2 = 1$ and (σ_t) is a nonnegative process obeying the recurrence equation

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i(|X_{t-i}| - \gamma X_{t-i})^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2\,, \qquad t \in \mathbb{Z}\,. \tag{3.9}$$

The parameters α_i and β_j are assumed nonnegative in order to ensure the nonnegativity of the squared volatility process (σ_t^2), and $|\gamma| \le 1$.

We will see in Remark 3.3.2 below that $\sqrt{\alpha_0}$ plays the role of a scale parameter. If $\gamma = 0$, then we have the so–called GARCH(p, q) model, which was introduced by Bollerslev [18] as an extension of ARCH earlier developed by Engle [46]. The ARCH model corresponds to $q = 0$. The AGARCH(p, q) models form a subset of the so–called asymmetric power GARCH models, which were proposed by Ding et al. [39]. Observe that $|\gamma| \le 1$ is a necessary identifiability condition. Indeed, if $|\gamma| > 1$, the parameters

$$\tilde{\gamma} = \frac{1}{\gamma}\,, \qquad \tilde{\alpha}_i = \frac{(1+\gamma)^2}{(1+1/\gamma)^2}\,\alpha_i\,, \qquad i \ge 1,$$

would lead to an identical model for the (squared) volatility since $\alpha_i(1 + \gamma)^2 = \tilde{\alpha}_i(1 + \tilde{\gamma})^2$ and $\alpha_i(1-\gamma)^2 = \tilde{\alpha}_1(1-\tilde{\gamma})^2$. Note that one can interpret AGARCH(p, q) as a special case of a so–called threshold GARCH(p, q) model since (3.9) may also be written as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i((1-\gamma)^2 + 4\gamma \mathbf{1}_{\{X_{t-i}<0\}})X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$$

Here $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function. Threshold GARCH models were independently introduced by Glosten et al. [57] and Zakoïan [135]. The latter representation also shows that the volatility responds asymmetrically to rises and falls in stock prices if $\gamma \neq 0$; the case where negative log–returns have a larger impact than positive log–returns corresponds to $\gamma > 0$ and the converse case to $\gamma < 0$. The AGARCH model allows leverage effects in the sense of Section 3.1.4.

GARCH (and some of its relatives) gained popularity in financial econometrics and among practitioners because it often gives a reasonable fit to real–life data and can explain some of the stylized facts of financial log–returns. There is nowadays a vast econometric literature supporting the relevance of GARCH, see e.g. Shephard [123] for a review of the literature until 1996. A nice example of a successful application of GARCH models in financial risk management is the thorough study by McNeil and Frey [95].

Stationarity

Despite the seemingly simple defining equations (3.8)–(3.9), the stochastic properties of AGARCH are not easy to deduce. The problem of finding a necessary and sufficient condition for stationarity of GARCH waited for a solution until Nelson [105] provided the answer for GARCH$(1,1)$ and Bougerol and Picard [21] for the general GARCH(p,q) case.

The main idea for tackling the stationarity question is to write the squared AGARCH(p,q) process in state space form and to analyze the resulting stochastic recurrence equation. Since an AGARCH(p,q) process is also an AGARCH(p',q') process for any $p' \geq p$, $q' \geq q$, we may without loss of generality assume $p,q \geq 2$. We then introduce the $(p+q-1)$–dimensional vectors

$$\mathbf{Y}_t = (\sigma_t^2, \ldots, \sigma_{t-q+1}^2, (|X_{t-1}| - \gamma X_{t-1})^2, \ldots, (|X_{t-p+1}| - \gamma X_{t-p+1})^2)^T,$$

$$\mathbf{B} = (\alpha_0, 0, \ldots, 0)^T,$$

and the $(p+q-1) \times (p+q-1)$ matrices

$$\mathbf{A}_t = \begin{pmatrix} \alpha_1(|Z_t| - \gamma Z_t)^2 + \beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q & \alpha_2 & \alpha_3 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ (|Z_t| - \gamma Z_t)^2 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \qquad t \in \mathbb{Z}. \tag{3.10}$$

Then the two equations (3.8)–(3.9), imply

$$\mathbf{Y}_{t+1} = \mathbf{A}_t \mathbf{Y}_t + \mathbf{B}, \qquad t \in \mathbb{Z}. \tag{3.11}$$

Observe that $((\mathbf{A}_t, \mathbf{B}))$ is an iid sequence. A functional relationship between the present and the past of a stochastic process is also called a stochastic recurrence equation (SRE); see Section 2.6 for a precise definition of the notion of a SRE. Therefore (3.11) is a linear SRE on the state space $\mathbb{R}^{p+q-1}$; the term "linear" is justified by the fact that the map which transforms $\mathbf{Y}_t$ into $\mathbf{Y}_{t+1}$ is (affine) linear. Alternatively, one could view (3.11) as a stochastic dynamical system or a vector autoregression with random coefficients $(\mathbf{A}_t)$. A solution to a given SRE is a stochastic process obeying the corresponding stipulated relationship.

It is not difficult to see that (3.8)–(3.9) have a unique stationary solution if and only if (3.11) has a unique stationary solution with nonnegative coordinates (for short: a unique nonnegative stationary solution). We only have to give an argument for the sufficiency of the latter statement since necessity has been shown already by derivation of (3.11). Assume that the linear SRE (3.11) has a unique nonnegative stationary solution. Then it can be verified that the first coordinate of $\mathbf{Y}_t$ obeys (3.9), and X_t is then obtained via relation (3.8).

Therefore the stationarity question has been translated into the problem of existence and uniqueness of a nonnegative solution to the linear SRE (3.11), where $\mathbf{A}_t$ and $\mathbf{B}$ are as above. A complete answer to the latter problem was given in Bougerol and Picard [21]. Roughly speaking, one has to show that the random map $\mathbf{y} \mapsto \mathbf{A}_0\mathbf{y} + \mathbf{B}$ is "contractive on average". More precisely, the so–called top Lyapunov exponent associated with $(\mathbf{A}_t)$ has to be negative, i.e.,

$$\rho = \inf_{t\in\mathbb{N}} \left\{ \frac{1}{t+1} \mathbb{E}(\log \|\mathbf{A}_0 \cdots \mathbf{A}_{-t}\|_{\text{op}}) \right\} < 0, \tag{3.12}$$

where $\|\cdot\|_{\text{op}}$ stands for the matrix operator norm corresponding to the Euclidean norm, see Section 2.3. Actually the value of the top Lyapunov exponent ρ is independent of the norm $|\cdot|$ because in finite–dimensional vector spaces all norms are equivalent and because ρ is also characterized through the limit relation (3.14) below; alternatively one might work with a non–Euclidean norm. We exclude the case $\alpha_0 = 0$ because it would lead to the trivial solution $X_t \equiv 0$. The following theorem is a straightforward generalization of the results by Bougerol and Picard [21] on the stationarity of GARCH(p,q).

Theorem 3.3.1. *Let $\alpha_0 > 0$ and α_i, β_j be nonnegative numbers and $|\gamma| \le 1$. Then the* AGARCH *equations* (3.8)–(3.9) *admit a unique stationary ergodic solution* $((X_t,\sigma_t))$ *if and only if the condition* (3.12) *holds. The corresponding solution* $(\mathbf{Y}_t)$ *of the* SRE (3.11) *has the almost sure representation*

$$\mathbf{Y}_t = \mathbf{B} + \sum_{j=1}^{\infty} \Big(\prod_{i=1}^{j} \mathbf{A}_{t-i} \Big) \mathbf{B}\,. \tag{3.13}$$

Remark 3.3.2. Note that $\mathbf{Y}_t$ is $\mathcal{F}_{t-1}$–measurable for every $t \in \mathbb{Z}$ and for this reason every stationary AGARCH(p,q) process belongs to the general conditionally heteroscedastic class given in (3.6). In other words, every stationary AGARCH(p,q) process is automatically causal, i.e., X_t is $\mathcal{F}_t$–measurable for every t. This is in contrast to ARMA processes, see Example 3.1.2 in Brockwell and Davis [29]. Observe also that σ_t is $\mathcal{F}_{t-1}$–measurable and thus independent of Z_t for every fixed $t \in \mathbb{Z}$. Since $(\mathbf{Y}_t)$ is homogeneous in α_0, the parameter $\sqrt{\alpha_0}$ is recognized as a scale parameter for $((X_t,\sigma_t))$. This entails: if (X_t) is an AGARCH(p,q) process with parameters $\boldsymbol{\theta}_0 = (\alpha_0,\dots,\alpha_p,\beta_1,\dots\beta_q,\gamma)^T$, then for any $\lambda > 0$ the process $(\sqrt{\lambda}X_t)$ is AGARCH(p,q) with parameters $\tilde{\boldsymbol{\theta}}_0 = (\lambda\alpha_0,\alpha_1,\dots,\alpha_p,\beta_1,\dots,\beta_q,\gamma)^T$ and identical innovations (Z_t). □

Proof. The sufficiency of $\rho < 0$ for the existence and uniqueness of a nonnegative solution to (3.11) follows from a direct application of Theorem 2.6.1. We rephrase the SRE (3.11) on $\mathbb{R}^{p+q-1}$ as

$$\mathbf{Y}_{t+1} = \phi_t(\mathbf{Y}_t), \qquad t \in \mathbb{Z},$$

with

$$\phi_t(\mathbf{y}) = \mathbf{A}_t \mathbf{y} + \mathbf{B}.$$

In order to check the assumptions of Theorem 2.6.1, we work with the metric induced by the Euclidean norm in $\mathbb{R}^{p+q-1}$. Then it is easily seen from a comparison of (2.9) and (2.12) that

$$\Lambda(\phi_0 \circ \cdots \circ \phi_{-r+1}) = \|\mathbf{A}_0 \cdots \mathbf{A}_{-r+1}\|_{\mathrm{op}}.$$

Thus (3.12) implies $\mathbb{E}(\log \Lambda(\phi_0^{(r)})) < 0$ for some $r \geq 1$, and condition (2.16) of Theorem 2.6.1 is met. Moreover, $\mathbb{E}[\log^+ \Lambda(\phi_0)] = \mathbb{E}(\log^+ \|\mathbf{A}_0\|_{\mathrm{op}}) < \infty$ trivially follows from $\mathbb{E}\|\mathbf{A}_0\|_{\mathrm{op}} < \infty$, which is a consequence of $\mathbb{E}(Z_0^2) = 1$ and the fact that every matrix norm is equivalent to the Frobenius norm (2.7). Condition S.1 of Theorem 2.6.1 is trivially true if we take $\mathbf{y} = \mathbf{0}$. The expression of the right–hand side of (3.13) is the limit $m \to \infty$ of the backward iterates $\phi_{t-1} \circ \cdots \circ \phi_{t-m}(\mathbf{0})$ and has nonnegative coordinates. We conclude the proof of the sufficiency part by mentioning that Bougerol and Picard [21] did not use the "high–level" Theorem 2.6.1 in order to establish the sufficiency of $\rho < 0$ (with $\gamma = 0$) for the stationarity of GARCH(p,q). They directly applied Kingman's subadditive ergodic theorem [74] to norms of products of random matrices, which yields

$$\frac{1}{j} \log \Big\| \prod_{i=1}^{j} \mathbf{A}_{t-i} \Big\|_{\mathrm{op}} \xrightarrow{\text{a.s.}} \rho < 0, \qquad j \to \infty. \tag{3.14}$$

Hence $\| \prod_{i=1}^{j} \mathbf{A}_{t-i} \|_{\mathrm{op}} \xrightarrow{\text{e.a.s.}} 0$, which shows that the series (3.13) is absolutely continuous a.s. Furthermore (3.13) defines a nonnegative stationary ergodic solution to (3.11). This method is not essentially different from the lines of proof of Theorem 2.6.1 presented in Section 2.6 of this monograph.

The proof that condition (3.12) is also necessary requires the use of results from the theory of random matrices. More or less we copy the arguments in Bougerol and Picard [21]. Suppose $(\mathbf{Y}_t)$ is a solution of (3.11). We then need to show that the top Lyapunov exponent ρ associated with $(\mathbf{A}_t)$ is negative. Using the iteration (3.11) repeatedly, we have

$$\begin{aligned} \mathbf{Y}_0 &= \mathbf{A}_{-1}\mathbf{Y}_{-1} + \mathbf{B} = \mathbf{A}_{-1}\mathbf{A}_{-2}\mathbf{Y}_{-2} + \mathbf{A}_{-1}\mathbf{B} + \mathbf{B} \\ &= \quad \cdots \quad = \mathbf{A}_{-1}\mathbf{A}_{-2}\cdots\mathbf{A}_{-t-1}\mathbf{Y}_{-t-1} + \mathbf{B} + \sum_{j=1}^{t} \Big(\prod_{i=1}^{j} \mathbf{A}_{-i} \Big) \mathbf{B}. \end{aligned}$$

Since all coefficients of $\mathbf{A}_t$, $\mathbf{B}_t$ and $\mathbf{Y}_t$ are nonnegative,

$$\sum_{j=1}^{t} \Big(\prod_{i=1}^{j} \mathbf{A}_{-i} \Big) \mathbf{B} \leq \mathbf{Y}_0,$$

where the relation "$\geq$" has to be understood componentwise. For this reason the series $\sum_{j=1}^{t}(\prod_{i=1}^{j} \mathbf{A}_{-i})\mathbf{B}$ converges a.s. To show that this implies $\rho < 0$ one has to make use of a result from the theory of random matrices. Observe that $(\mathbf{A}_t)$ is iid with $\mathbb{E}(\log^+ \|\mathbf{A}_0\|_{\text{op}}) < \infty$. Then Lemma 2.1 in Bougerol and Picard [21] says that $\rho < 0$ is implied by

$$\lim_{t\to\infty} \|\mathbf{A}_0 \cdots \mathbf{A}_{-t}\|_{\text{op}} = 0 \quad \text{a.s.} \tag{3.15}$$

If $(\mathbf{e}_k)$ denotes the canonical basis in $\mathbb{R}^{p+q-1}$, for (3.15) it is enough to prove that

$$\lim_{t\to\infty} \mathbf{A}_0 \cdots \mathbf{A}_{-t}\mathbf{e}_k = \mathbf{0} \quad \text{a.s.} \tag{3.16}$$

for each $1 \leq k \leq p + q - 1$. Since $\mathbf{B} = \alpha_0\mathbf{e}_1$, $\alpha_0 > 0$ and the series $\sum_{j=1}^{t}(\prod_{i=1}^{j} \mathbf{A}_{-i})\mathbf{B}$ converges a.s., relation (3.16) is certainly true for $k = 1$. Furthermore, $\mathbf{A}_{-t}\mathbf{e}_q = \beta_p\mathbf{e}_1$ together with (3.16) for $k = 1$ imply (3.16) for $k = q$. For the indices $k = (q-1), \ldots, 2$, the property (3.16) follows from a backward recursion. Indeed, if (3.16) is valid for $3 \leq k \leq q$, then $\mathbf{A}_{-t}\mathbf{e}_k = \beta_k\mathbf{e}_1 + \mathbf{e}_{k+1}$, implies

$$\lim_{t\to\infty} \mathbf{A}_0 \cdots \mathbf{A}_{-t}\mathbf{e}_k = \beta_k \lim_{t\to\infty} \mathbf{A}_0 \cdots \mathbf{A}_{-t+1}\mathbf{e}_1 + \lim_{t\to\infty} \mathbf{A}_0 \cdots \mathbf{A}_{-t+1}\mathbf{e}_{k+1} = \mathbf{0}.$$

Using the same arguments for $\mathbf{e}_{q+1}, \ldots, \mathbf{e}_{p+q-1}$ and

$$\mathbf{A}_{-t}\mathbf{e}_{p+q-1} = \alpha_p\mathbf{e}_1, \qquad \mathbf{A}_{-t}\mathbf{e}_{q+k-1} = \alpha_k\mathbf{e}_1 + \mathbf{e}_{q+k}$$

for $2 \leq k \leq p-1$, we see that (3.16) holds for each $\mathbf{e}_k$, and hence $\rho < 0$. This completes the proof. □

Given coefficients α_i and β_j and the distribution of Z_0, the verification of condition (3.12) in general requires a simulation approach based on (3.14) because there are no tractable expressions for the norms of matrix products $\mathbf{A}_0 \cdots \mathbf{A}_{-t}$. There exist however special cases, for which one can easily decide whether the top Lyapunov exponent $\rho < 0$ or $\rho \geq 0$.

Proposition 3.3.3. *The following statements hold true for the top Lyapunov exponent ρ associated with $(\mathbf{A}_t)$.*

(a) *If $\mathbb{E}(\log \|\mathbf{A}_0 \circ \mathbf{A}_{-r+1}\|_{\text{op}}) < 0$ for some $r \geq 1$, then $\rho < 0$.*

(b) *The relation*

$$\sum_{i=1}^{p} \alpha_i \mathbb{E}[(|Z_0| - \gamma Z_0)^2] + \sum_{j=1}^{q} \beta_j < 1$$

implies $\rho < 0$.

(c) *If the parameters fulfill* $\alpha_i, \beta_j > 0$ *for all* i, j *and* $|\gamma| < 1$ *and the innovations distribution is such that* $\mathbb{P}(|Z_0| = 0) = 0$ *and* $\mathbb{P}(|Z_0| \le z) < 1$ *for all* $z \ge 0$*, then*

$$\sum_{i=1}^{p} \alpha_i \mathbb{E}[(|Z_0| - \gamma Z_0)^2] + \sum_{j=1}^{q} \beta_j = 1$$

implies $\rho < 0$.

(d) *If* $\sum_{j=1}^{q} \beta_j \ge 1$ *then* $\rho \ge 0$.

Proof. (a): This implication is obvious from the definition (3.12) of the top Lyapunov exponent ρ. (b): Arguing by induction on p and expanding the determinant with respect to the last column, one can verify that $\mathbb{E}\mathbf{A}_0$ has characteristic polynomial

$$\det(\lambda \mathbf{I}_{p+q-1} - \mathbb{E}\mathbf{A}_0) = \lambda^{p+q-1}\Big(1 - \sum_{i=1}^{p} \alpha_i \mathbb{E}[(|Z_0| - \gamma Z_0)^2]\lambda^{-i} - \sum_{j=1}^{q} \beta_j \lambda^{-j}\Big).$$

Here $\mathbf{I}_{p+q-1}$ denotes the identity matrix of dimension $p+q-1$. If $|\lambda| \ge 1$, then by the triangle inequality $|x - y| \ge |x| - |y|$ applied to the latter equation,

$$\det(\lambda \mathbf{I}_{p+q-1} - \mathbb{E}\mathbf{A}_0) \ge 1 - \sum_{i=1}^{p} \alpha_i \mathbb{E}[(|Z_0| - \gamma Z_0)^2] - \sum_{j=1}^{q} \beta_j > 0,$$

which shows that $\mathbb{E}\mathbf{A}_0$ has spectral radius $\varrho < 1$. Recall that the spectral radius of a matrix is its maximal absolute eigenvalue. Since the top Lyapunov exponent always fulfills $\rho \le \log \varrho$ (see (1.4) in Kesten and Spitzer [73]) the assertion is proved. An alternative proof will be provided in Proposition 3.3.5 below.
(c), (d): The proofs of these statements again parallel those of Corollaries 2.2 and 2.3 in Bougerol and Picard [21] and so we omit the details. □

Stationarity of AGARCH(1,1)

If $p = q = 1$, i.e., in the special case of AGARCH$(1,1)$, it is enough to consider the linear SRE

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1(|X_t| - \gamma X_t)^2 + \beta_1 \sigma_t^2 = A_t \sigma_t^2 + \alpha_0 \tag{3.17}$$

with $A_t = \alpha_1(|Z_t| - \gamma Z_t)^2 + \beta_1$. Notice that (A_t) constitutes an iid sequence. The latter SRE has a unique nonnegative stationary solution if and only if

$$\mathbb{E}(\log A_0) = \mathbb{E}[\log(\alpha_1(|Z_0| - \gamma Z_0)^2 + \beta_1)] < 0, \tag{3.18}$$

and in that case, similarly to (3.13), the almost sure representation

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} \Big(\prod_{i=1}^{j} A_{t-i} \Big)$$
$$= \alpha_0 \Big(1 + \sum_{j=1}^{\infty} \prod_{i=1}^{j} (\alpha_1 (|Z_{t-i}| - \gamma Z_{t-i})^2 + \beta_1) \Big) \tag{3.19}$$

is valid. The sufficiency of (3.18) for stationarity is a consequence of the fact that the map $s \mapsto \psi_0(s) = A_0 s + \alpha_0$ has Lipschitz coefficient $\Lambda(\psi_0) = A_0$ together with an application of Theorem 2.6.1. Alternatively one may directly check via Lemma 2.5.5 that $\prod_{i=1}^{j} A_{t-i} \xrightarrow{\text{e.a.s.}} 0$ as $j \to \infty$. This then implies that the series in (3.19) converges absolutely a.s., and then it is readily verified that (3.19) provides a solution of (3.17). If $(\tilde{\sigma}_t^2)$ is yet another stationary solution of (3.17), then since (σ_t^2) and $(\tilde{\sigma}_t^2)$ are stationary and $\prod_{i=1}^{j} A_{t-i} \xrightarrow{\text{e.a.s.}} 0$ one has that

$$|\sigma_t^2 - \tilde{\sigma}_t^2| = A_{t-1} \cdots A_{t-j} |\sigma_{t-j}^2 - \tilde{\sigma}_{t-j}^2|$$
$$\leq A_{t-1} \cdots A_{t-j} (\sigma_{t-j}^2 + \tilde{\sigma}_{t-j}^2) \xrightarrow{\mathbb{P}} 0, \qquad j \to \infty$$

for every t. This shows $\mathbb{P}(\tilde{\sigma}_t^2 = \sigma_t^2) = 1$, i.e., the uniqueness of the solution (σ_t^2). The stationarity and ergodicity of (3.19) is clear from Proposition 2.1.1.

An argument for the necessity of condition (3.18) for the stationarity of AGARCH$(1,1)$ can be given as follows. If (σ_t^2) is stationary and obeys (3.17), then by repeated application of (3.17) together with the nonnegativity of A_t and α_0,

$$\sigma_t^2 \geq \alpha_0 \Big(1 + \sum_{j=1}^{\infty} \Big(\prod_{i=1}^{j} A_{t-i} \Big) \Big) \tag{3.20}$$

Assuming $\mathbb{E}(\log A_0) \geq 0$ by contradiction, we have

$$\limsup_{j \to \infty} \sum_{i=1}^{j} \log(A_{t-i}) = \infty \quad \text{a.s.}$$

since $(\sum_{i=1}^{j} \log(A_{t-i}))_{j \geq 1}$ is a random walk with nonnegative drift $\mathbb{E}(\log A_0)$. As a consequence of this limit result applied to the lower bound (3.20) and $A_t \geq 0$ for every t, we get $\sigma_t^2 = \infty$ a.s. This is the desired contradiction and hence necessarily $\mathbb{E}(\log A_0) < 0$.

The arguments we have given here go back to Nelson [106]. If $\gamma = 0$, the condition (3.18) reduces to

$$\mathbb{E}[\log(\alpha_1 Z_0^2 + \beta_1)] < 0,$$

which is Nelson's sufficient and necessary condition for stationarity in the GARCH$(1,1)$ model. Summarizing, we have treated the stationarity in the

AGARCH$(1,1)$ process by means of elementary techniques and have obtained a rather tractable necessary and sufficient criterion for stationarity as compared to Theorem 3.3.1. An important conclusion from (3.18) is the insight that there *exist* stationary AGARCH$(1,1)$ processes with $\alpha_1+\beta_1>1$. Indeed, if $\alpha_1+\beta_1=1$ and $\gamma=0$, then by Jensen's inequality

$$\mathbb{E}[\log(\alpha_1 Z_0^2+\beta_1)]<\log(\mathbb{E}[\alpha_1 Z_0^2+\beta_1])=\log 1=0,$$

which entails stationarity. Now a continuity argument indicates that there are $\alpha_1+\beta_1>1$ with $\mathbb{E}[\log(\alpha_1 Z_0^2+\beta_1)]<0$.

Heavy Tails

Under mild conditions on the distribution of Z_0, the unconditional distribution of a stationary AGARCH process is heavy–tailed. We restrict ourselves to AGARCH$(1,1)$, because only the results of this case will be used later on. We now give a heuristic argument for the heavy–tailedness. Recall that the squared volatility of an AGARCH$(1,1)$ process obeys the linear SRE

$$\sigma_{t+1}^2=A_t\sigma_t^2+\alpha_0, \qquad t\in\mathbb{Z}, \tag{3.21}$$

where $A_t=\alpha_1(|Z_t|-\gamma Z_t)^2+\beta_1$. Suppose $\mathbb{E}A_0^{q/2}\ge 1$ for some $q>0$. Then, since A_t and σ_{t-1} are nonnegative and $\alpha_0>0$, equation (3.21) implies

$$\sigma_{t+1}^q>A_t^{q/2}\sigma_t^q.$$

Taking the expectation on both sides of the latter inequality, using that A_t and σ_t^2 are independent and exploiting the stationarity of (A_t) and (σ_t^2), one may conclude that

$$\mathbb{E}\sigma_0^q>\mathbb{E}A_0^{q/2}\,\mathbb{E}\sigma_0^q\ge\mathbb{E}\sigma_0^q.$$

The latter inequality implies $\mathbb{E}\sigma_0^q=\infty$. On the other hand, an application of the Minkowski inequality to the representation (3.19) yields for any $q>0$ that

$$\mathbb{E}\sigma_0^q\le\alpha_0^{q/2}\Big(1+\sum_{j=1}^{\infty}(\mathbb{E}A_0^{q/2})^{j/\max(1,q/2)}\Big)^{\max(1,q/2)},$$

which is finite if $\mathbb{E}A_0^{q/2}<1$. Altogether, the following criterion is valid:

$$\mathbb{E}|X_0|^q,\,\mathbb{E}\sigma_0^q<\infty \iff \mathbb{E}A_0^{q/2}<1, \qquad q>0.$$

Note that under the mild assumption $\mathbb{P}(A_0>1)>0$ there exists a $q>0$ with $\mathbb{E}A_0^{q/2}\ge 1$ since $\mathbb{E}A_0^s\to\infty$ as $s\to\infty$. Hence the random variables $|X_0|$ and σ_0 do not have finite moments of all orders; hence they are heavy–tailed. This statement can be made more precise. Under regularity conditions, one can show that the existence of a positive solution κ to the equation $\mathbb{E}A_0^{s/2}=1$

implies that $\mathbb{P}(\sigma_0 > x) \sim cx^{-\kappa}$ as $x \to \infty$. Conditions for the validity of this statement were given in Mikosch and Stărică [99] for GARCH(1, 1) and in Basrak et al. [5] for GARCH(p, q). They are an immediate consequence of work by Kesten [72] and Goldie [58] on the tail behavior of stationary solutions to linear SREs.

In the next theorem we present the version for AGARCH(1, 1), which is a straightforward generalization of the results by Mikosch and Stărică [99]. Notice the remarkable phenomenon that the distribution of X_t *is* heavy–tailed even if Z_t has *light* tails, e.g. if $Z_t \sim \mathcal{N}(0, 1)$.

Theorem 3.3.4. *Assume that the parameters $\alpha_1, \beta_1, \gamma$ and the distribution of Z_0 satisfy the following conditions:*

(i) *Z_0 has a positive density on $\mathbb{R}$.*

(ii) *Let the parameters fulfill $\alpha_0 > 0$, $\alpha_1, \beta_1 \geq 0$ and $|\gamma| \leq 1$ and suppose that $\mathbb{E}(\log A_0) = \mathbb{E}[\log(\alpha_1(|Z_0| - \gamma Z_0)^2 + \beta_1)] < 0$ (so that there exists a unique stationary ergodic* AGARCH(1, 1) *process, which is nontrivial).*

(iii) *There exists $s_0 < \infty$ such that $\mathbb{E}A_0^s < \infty$ for all $s < s_0$ and $\mathbb{E}A_0^{s_0} = \infty$ or $\mathbb{E}A_0^{s_0} < \infty$ for all $s \geq 0$ and $\lim_{s\to\infty} \mathbb{E}A_0^s = \infty$.*

Then the following statements hold:

(A) *The equation*

$$\mathbb{E}A_0^{s/2} = 1$$

has a unique positive solution κ.

(B) *The unique stationary solution $((X_t, \sigma_t))$ to the* AGARCH(1, 1) *equations* (3.8)–(3.9) *satisfies*

$$\mathbb{P}(|X_0| > x) \sim \mathbb{E}|Z_0|^\kappa \, \mathbb{P}(\sigma_0 > x) \sim c_0 \, x^{-\kappa}, \qquad x \to \infty, \tag{3.22}$$

for some $c_0 > 0$.

In what follows, we refer to κ as the tail index.

Basrak et al. [5] have put up precise conditions under which the tails of the distribution function of $|X_0|$ and σ_0 in GARCH(p, q) are Pareto–like in the sense that (3.22) holds for certain $\kappa, c_0 > 0$. However, the characterization of the tail index κ is less explicit compared to (A). An extension of these results to AGARCH(p, q) would be possible.

To conclude this section, we mention a useful criterion for a finite unconditional variance in AGARCH(p, q), or in other words, for covariance stationarity (or weak stationarity).

Proposition 3.3.5. *Let (X_t) be a stationary* AGARCH(p, q) *process with $\alpha_0 > 0$. Then if*

$$\sum_{i=1}^{p} \alpha_i \mathbb{E}[(|Z_0| - \gamma Z_0)^2] + \sum_{j=1}^{q} \beta_j < 1, \tag{3.23}$$

the random variable X_0 has finite variance equal to

$$\mathbb{E}X_0^2 = \alpha_0 \Big(1 - \sum_{i=1}^{p} \alpha_i \mathbb{E}[(|Z_0| - \gamma Z_0)^2] - \sum_{j=1}^{q} \beta_j\Big)^{-1}. \tag{3.24}$$

Otherwise $\mathbb{E}X_0^2 = \infty$.

Proof. In the GARCH(p, q) case (i.e. $\gamma = 0$), this criterion was given by Bollerslev [18]. Necessity is relatively easy to obtain. Since $X_t = \sigma_t Z_t$ and σ_t and Z_t are independent, $\mathbb{E}X_0^2 = \mathbb{E}\sigma_0^2$. Taking the expectation on both sides of

$$\sigma_0^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i (|Z_{-i}| - \gamma Z_{-i})^2 \sigma_{-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{-j}^2,$$

and accounting for the stationarity of (σ_t^2) and for the independence of σ_{-i} and Z_{-i}, we receive

$$\mathbb{E}\sigma_0^2 = \alpha_0 + \mathbb{E}\sigma_0^2 \left(\sum_{i=1}^{p} \alpha_i \mathbb{E}[(|Z_{-i}| - \gamma Z_{-i})^2] + \sum_{j=1}^{q} \beta_j \right).$$

This equation in $\mathbb{E}\sigma_0^2$ has a finite nonnegative solution if and only if relation (3.23) holds.

Regarding the sufficiency of (3.23) for (3.24), one first derives a Volterra type series expansion of σ_t^2. We mention that the same approach was taken by Giraitis et al. [55] in their analysis of (covariance) stationarity of ARCH(∞) processes. Since every AGARCH(p, q) process can be written as an AGARCH$(\max(p, q), \max(p, q))$ process, we may assume $p = q$ without loss of generality. Introduce

$$C_t^{(i)} = \alpha_i (|Z_t| - \gamma Z_t)^2 + \beta_i, \qquad t \in \mathbb{Z}$$

for $i \in [1, p]$ and note that (3.9) with $p = q$ then reads

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} C_{t-i}^{(i)} \sigma_{t-i}^2. \tag{3.25}$$

Subsequently replacing σ_{t-i}^2 by the equivalent expression $\alpha_0 + \sum_{j=1}^{p} C_{t-i-j}^{(i)} \sigma_{t-i-j}^2$ in (3.25), we formally obtain

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \sum_{i_1=1}^{p} C_{t-i_1}^{(i_1)} \sigma_{t-i_1}^2 \\
&= \alpha_0 + \alpha_0 \sum_{i_1=1}^{p} C_{t-i_1}^{(i_1)} + \sum_{i_1=1}^{p} \sum_{i_2=1}^{p} C_{t-i_1}^{(i_1)} C_{t-i_1-i_2}^{(i_2)} \sigma_{t-i_1-i_2}^2 \\
&\vdots \qquad\quad \vdots \\
&= \alpha_0 \Big(1 + \sum_{k=1}^{\infty} M_t(k)\Big), \qquad (3.26)
\end{aligned}$$

where

$$M_t(k) = \sum_{i_1,\ldots,i_k=1}^{p} C_{t-i_1}^{(i_1)} C_{t-i_1-i_2}^{(i_2)} \cdots C_{t-i_1-\cdots-i_k}^{(i_k)}. \qquad (3.27)$$

Now our task is to show that the Volterra series expansion (3.26) of σ_t^2 is valid under (3.23). To do so, it is enough to verify that $\sum_{k=1}^{\infty} M_t(k) < \infty$ a.s. Indeed, since the right–hand side of (3.26) obeys (3.25) and since the AGARCH equations have a unique solution (Theorem 3.3.1), the representation (3.26) is clearly true. In order to establish $\sum_{k=1}^{\infty} M_t(k) < \infty$ a.s., we compute the first moment of $M_t(k)$. Using the fact that the factors appearing under the sum (3.27) are independent for each fixed k, we have

$$\mathbb{E}[M_t(1)] = \sum_{i=1}^{p} \mathbb{E}C_0^{(i)} = \sum_{i=1}^{p} (\alpha_i \mathbb{E}[(|Z_0| - \gamma Z_0)^2] + \beta_i) =: \lambda. \qquad (3.28)$$

Moreover, since

$$M_t(k+1) = \sum_{i_1=1}^{p} C_{t-i_1}^{(i_1)} M_{t-i_1}(k),$$

since the random variables $C_{t-i_1}^{(i_1)}$ and $M_{t-i_1}(k)$ are independent for every i_1 and since $\mathbb{E}[M_t(k)]$ does not depend on t,

$$\mathbb{E}[M_t(k+1)] = \lambda \mathbb{E}[M_t(k)].$$

Combining this with (3.28) we finally obtain

$$\mathbb{E}[M_t(k)] = \lambda^k, \qquad k \geq 1.$$

Recall that $\lambda < 1$ by assumption (3.23). Now an application of Fubini's theorem gives $\mathbb{E}[\sum_{k=1}^{\infty} M_t(k)] = \sum_{k=1}^{\infty} \lambda^k < \infty$ and shows $\sum_{k=1}^{\infty} M_t(k) < \infty$ a.s. As a by–product we receive $\mathbb{E}\sigma_t^2 < \infty$, which completes the proof. □

IGARCH

Let (X_t) be a stationary GARCH(p,q) process with $\alpha_0 > 0$. Then the boundary case of parameters

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1 \tag{3.29}$$

is referred to as integrated GARCH (or IGARCH); here the term "integrated" stems from the fact that (X_t^2) is *formally* an ARMA process with an AR part containing a unit root. We refer to Section 4.2.2 for a more detailed analysis of the squared GARCH process. From a straightforward application of Theorem 3.3.4 (with $\gamma = 0$) it follows that the IGARCH$(1,1)$ process has a marginal distribution with tail index $\kappa = 2$, provided the regularity assumptions (i)–(iii) of Proposition 3.3.4 hold. To the best of our knowledge, "integrated AGARCH processes" have not been introduced in the literature.

3.3.2 EGARCH Models

In its simplest form, the exponential GARCH (EGARCH) process of Nelson [106] is of form

$$X_t = \sigma_t Z_t, \tag{3.30}$$

with a squared volatility obeying

$$\log \sigma_t^2 = \alpha + \beta \log \sigma_{t-1}^2 + \gamma Z_{t-1} + \delta |Z_{t-1}|, \qquad t \in \mathbb{Z}, \tag{3.31}$$

where $\alpha, \gamma, \delta \in \mathbb{R}$ and $|\beta| < 1$. Note that the sequence $(\log \sigma_t^2)$ constitutes a causal AR(1) process with mean $\mu = (\alpha + \delta \mathbb{E}|Z_0|)/(1-\beta)$ and error sequence $(\gamma Z_{t-1} + \delta(|Z_{t-1}| - \mathbb{E}|Z_0|))$. Since the innovations have a finite mean and $|\beta| < 1$, it follows by the theory of ARMA processes that the unique stationary solution to (3.31) is given by

$$\log \sigma_t^2 = \alpha(1-\beta)^{-1} + \sum_{k=0}^{\infty} \beta^k \left(\gamma Z_{t-1-k} + \delta |Z_{t-1-k}|\right), \qquad t \in \mathbb{Z}. \tag{3.32}$$

Hence the stationarity issue does not pose any problems.

In addition we suppose that $0 \le \beta < 1$ and $\delta \ge |\gamma|$. This seems reasonable from an economics point of view. One expects a positive relationship between volatilities on successive days, i.e., $\beta \ge 0$. The squared volatility σ_t^2 as a function of Z_{t-1} should be nondecreasing on the positive real line (i.e. $\gamma + \delta \ge 0$) and non–increasing on the negative real line (i.e. $\delta - \gamma \ge 0$). Observe that the volatility responds asymmetrically to rises and falls in stock prices if and only if $\gamma \ne 0$; the case when negative log–returns have a larger impact than positive log–returns corresponds to $\gamma < 0$ and vice versa. Thus EGARCH captures leverage effects.. Altogether, $\gamma z + \delta |z| \ge 0$ for all $z \in \mathbb{R}$. From this we also deduce

$$\log \sigma_t^2 \geq \alpha(1-\beta)^{-1} \qquad \text{a.s.}$$

The tail behavior of an infinite series as it appears on the right–hand side of (3.32) is well–studied. Let us define nonnegative random variable $W_0 = (\gamma Z_0 + \delta|Z_0|)$ and let $M_{W_0}(r) = \mathbb{E}[e^{rW_0}]$ be its moment generating function. We start with the case that W_0 has a subexponential distribution; see Embrechts et al. [45] for the definition of subexponentiality. The subexponential class contains heavy–tailed distributions such as the lognormal or distributions with a tail which decays like a power function. A subexponentially distributed W_0 is not particularly interesting for practical applications because by the inequality $\log \sigma_1^2 \geq \alpha(1-\beta)^{-1} + W_0$ and the fact that $M_{W_0}(r) = \infty$, we would have $M_{\log \sigma_0^2}(r) = \infty$, which implies that any moment of σ_0 is infinite, i.e., $\mathbb{E}\sigma_0^q = \infty$ for all $q > 0$. This means that σ_0 would be extremely heavy–tailed.

In the case that W_0 has a distribution which admits a finite moment generating function, it is relatively simple to determine the moment generating function of $\log \sigma_0^2$. Indeed, by straightforward computation, which exploits the independence of the innovations $(\gamma Z_t + \delta|Z_t|)$, one shows that

$$\begin{aligned} M_{\log \sigma_0^2}(r) &= \exp(r\alpha(1-\beta)^{-1}) \prod_{k=0}^{\infty} M_{W_0}(\beta^k r) \\ &= \exp(r\alpha(1-\beta)^{-1}) \exp\Big(\sum_{k=0}^{\infty} \log M_{W_0}(\beta^k r)\Big). \end{aligned}$$

Now we can conclude that

$$\mathbb{E}\sigma_0^q = M_{\log \sigma_0^2}(q/2) = \exp(q\alpha(1-\beta)^{-1}/2) \exp\Big(\sum_{k=0}^{\infty} \log M_{W_0}(\beta^k q/2)\Big),$$

and thus obtain the criterion

$$\mathbb{E}\sigma_0^q < \infty \quad \iff \quad \sum_{k=0}^{\infty} \log M_{W_0}(\beta^k q/2) < \infty, \qquad q > 0.$$

There exist Abel–Tauber theorems relating tails and moment generating function in a neighborhood of zero, see Bingham et al. [14], but we do not develop the theory further. We also refer to recent work by Lindner and Meyer [87], who study the extremes of (X_t) in the special case of $\beta = 0$ and $Z_0 \sim \mathcal{N}(0,1)$.

3.4 Stochastic Volatility Models

For the sake of completeness we also mention the stochastic volatility models. A stochastic volatility model has form

$$X_t = \sigma_t Z_t, \qquad t \in \mathbb{Z},$$

where the sequences (σ_t) and (Z_t) are assumed independent of each other. Note that the conditionally heteroscedastic time series models we have seen so far, are *not* stochastic volatility models in the sense of the latter definition (although their volatility (σ_t) is a random process). Since in a stochastic volatility model (σ_t) and (Z_t) are independent of each other, (X_t) is stationary if and only if (σ_t) is stationary. A simple example is the case where $(\log \sigma_t^2)$ is modeled as a causal ARMA$(1,1)$ process independent of (Z_t), i.e.,

$$\log \sigma_t^2 = \alpha + \beta \log \sigma_{t-1}^2 + \eta_t, \qquad t \in \mathbb{Z}, \tag{3.33}$$

where α and $|\beta| < 1$ are the ARMA parameters and (η_t) is an iid sequence independent of (Z_t). Stochastic volatility models are usually not easy to handle statistically because it is rather difficult to estimate the unobserved volatilities (σ_t). Since we do not pursue stochastic volatility models furthermore, we merely mention the two comprehensive survey articles by Ghysels et al. [53] and Shephard [123], which also contain extensive bibliographies.

4

Parameter Estimation: An Overview

After a survey of estimation in ARMA models, we introduce the (Gaussian) quasi maximum likelihood estimator (QMLE) for the parameters of GARCH(p, q) and the so–called Whittle estimator, which is applied to the squares of a GARCH process. We summarize the existing results concerning the asymptotic behavior of these estimators and try to avoid technicalities. The emphasis is on general ideas and principles; this chapter is intended to facilitate the reading of Chapters 5 – 8, which are rather technical.

4.1 Estimation for ARMA Processes

In this section we give a short review on the estimation of ARMA parameters. We present three classical estimators: the Gaussian quasi maximum likelihood estimator, the least–squares estimator and the Whittle estimator. It has been well–known since the seminal paper by Hannan [64] that these three estimators are asymptotically equivalent, i.e., they have identical asymptotic behavior. The theory of estimation in linear ARMA models is an indispensable basis for a deeper understanding of the estimation problem in GARCH or other conditionally heteroscedastic time series models. We maintain the notation of Section 3.2 as far as possible.

4.1.1 Gaussian Quasi Maximum Likelihood Estimation

Suppose that (X_t) is a causal invertible ARMA(p, q) process given as the stationary solution to the difference equation

$$\varphi^\circ(B)X_t = \vartheta^\circ(B)Z_t,$$

where (Z_t) is iid with mean zero and finite variance $(\sigma^\circ)^2$ and where the characteristic polynomials

$$\varphi^\circ(z) = 1 - \varphi_1^\circ z - \cdots - \varphi_p^\circ z^p \quad \text{and} \quad \vartheta^\circ(z) = 1 + \vartheta_1^\circ z + \cdots + \vartheta_q^\circ z^q$$

are assumed to have no common zeros and $(\varphi_p^\circ, \vartheta_q^\circ) \neq (0,0)$. We want to estimate the true parameters $\boldsymbol{\theta}_0 = (\varphi_1^\circ, \ldots, \varphi_p^\circ, \vartheta_1^\circ, \ldots, \vartheta_q^\circ)^T$ and $(\sigma^\circ)^2$ from data $X_1, \ldots, X_n$. Observe that the type of the distribution of Z_t is *not* specified in the present framework, and hence it is impossible to determine a likelihood. A common approach in such a situation is to suppose that Z_t iid $\sim \mathcal{N}(0,1)$. Under this synthetic assumption, (X_t) is a zero–mean Gaussian process, and therefore it is possible to determine a so–called (Gaussian) quasi log–likelihood

$$\tilde{L}_n(\boldsymbol{\theta}, \sigma^2) = \log f_{\boldsymbol{\theta},\sigma^2}(X_1) + \sum_{t=2}^{n} \log f_{\boldsymbol{\theta},\sigma^2}(X_t \,|X_{t-1}, \ldots, X_1)\,, \tag{4.1}$$

where the symbol $f_{\boldsymbol{\theta},\sigma^2}(\cdot)$ stands for any kind of conditional or unconditional density related to a zero–mean Gaussian ARMA(p,q) process with parameters $\boldsymbol{\theta} = (\varphi_1, \ldots, \varphi_p, \vartheta_1, \ldots, \vartheta_q)^T$ and $\mathrm{Var}(Z_0) = \sigma^2$. Since the assumption that Z_t iid $\sim \mathcal{N}(0, \sigma^2)$ together with the parameter hypothesis $\boldsymbol{\theta}$ implies that the vector $(X_1, \ldots, X_t)^T$ is multivariate normally distributed (with a covariance matrix depending on $\boldsymbol{\theta}$ and σ^2), we have

$$X_t|X_{t-1}, \ldots, X_1 \;\sim\; \mathcal{N}(\tilde{X}_t(\boldsymbol{\theta}), \sigma^2 \tilde{r}_t(\boldsymbol{\theta})), \tag{4.2}$$

where

$$\tilde{X}_t(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta},\sigma^2}(X_t|X_{t-1}, \ldots, X_1) \quad \text{ and } \quad \tilde{r}_t(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta},\sigma^2}[(X_t - \tilde{X}_t(\boldsymbol{\theta}))^2]/\sigma^2.$$

Note that we used that $\mathrm{Var}(X_t \mid X_{t-1}, \ldots, X_1)$ is constant and consequently equal to $\tilde{r}_t(\boldsymbol{\theta})$ by virtue of the multivariate normality. Also observe that $\tilde{X}_t(\boldsymbol{\theta})$ coincides with the best linear predictor of X_t based on $X_1, \ldots, X_{t-1}$, and $\tilde{r}_t(\boldsymbol{\theta})$ is its normalized mean square error. We set $\tilde{X}_1(\boldsymbol{\theta}) = 0$ so that $\tilde{r}_1(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta},\sigma^2}[X_1^2]/\sigma^2$. In practice, the quantities $\tilde{X}_t(\boldsymbol{\theta})$ and $\tilde{r}_t(\boldsymbol{\theta})$ are determined by means of the so–called innovations algorithm, see Brockwell and Davis [29] . Combining (4.2) with (4.1) and taking logarithms, one obtains

$$\tilde{L}_n(\boldsymbol{\theta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left(\log(\sigma^2 \tilde{r}_t(\boldsymbol{\theta})) + \frac{(X_t - \tilde{X}_t(\boldsymbol{\theta}))^2}{\sigma^2 \tilde{r}_t(\boldsymbol{\theta})}\right). \tag{4.3}$$

The quasi maximum likelihood estimator (QMLE) is a maximizer $\binom{\tilde{\boldsymbol{\theta}}_n}{\tilde{\sigma}_n^2}$ of $\tilde{L}_n$ with respect to $\sigma^2 \in (0, \infty)$ and $\boldsymbol{\theta} \in \bar{C}$, where

$$C = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{p+q} \;\middle|\; \begin{array}{c} \varphi(z)\vartheta(z) \neq 0 \text{ for } |z| \leq 1, \\ \varphi(z), \vartheta(z) \text{ have no common zeros and } (\varphi_p, \vartheta_q) \neq (0,0) \end{array} \right\}.$$

In the definition of the latter set, the conditions that the characteristic polynomials $\varphi(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p$ and $\vartheta(z) = 1 + \vartheta_1 z + \cdots + \vartheta_q z^q$ must not have common zeros and $(\varphi_p, \vartheta_q) \neq (0,0)$ are necessary for the identifiability of the ARMA models. Identifiability means that there are no *distinct* vectors

$\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}} \in C$ which lead to the identical model. We also mention that by Remark 3 of Section 10.8 in Brockwell and Davis [29], the set C is open and bounded. Hannan [64] established the consistency and asymptotic normality of $\tilde{\boldsymbol{\theta}}_n$ in full generality. See also Brockwell and Davis [29] for a textbook treatment of the original Hannan proof.

Theorem 4.1.1 (Hannan [64]). *Let (X_t) be a causal invertible* ARMA(p,q) *process with true parameters $\boldsymbol{\theta}_0 \in C$ and $0 < \mathrm{Var}(Z_0) = (\sigma^\circ)^2 < \infty$. Then*

$$\tilde{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0, \qquad \tilde{\sigma}_n^2 \xrightarrow{\text{a.s.}} (\sigma^\circ)^2, \qquad n \to \infty,$$

and

$$\sqrt{n}\,(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, [\mathbf{W}(\boldsymbol{\theta}_0)]^{-1}), \qquad n \to \infty,$$

where the asymptotic covariance matrix is the inverse of

$$\mathbf{W}(\boldsymbol{\theta}_0) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\partial \log g(\lambda; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right)^T \left(\frac{\partial \log g(\lambda; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right) \mathrm{d}\lambda,$$

with

$$g(\lambda; \boldsymbol{\theta}) = \frac{|\vartheta(e^{-i\lambda})|^2}{|\varphi(e^{-i\lambda})|^2}, \qquad \lambda \in (-\pi, \pi].$$

Remark 4.1.2. Observe that the asymptotic covariance matrix $[\mathbf{W}(\boldsymbol{\theta}_0)]^{-1}$ of Theorem 4.1.1 does neither depend on the distribution nor on the innovations variance $(\sigma^\circ)^2$. For practical computations there exists an alternative representation of $\mathbf{W}(\boldsymbol{\theta}_0)$, see formula (8.8.3) in Brockwell and Davis [29]. □

Remark 4.1.3. Elementary calculus shows that $\tilde{\sigma}_n^2$ is related to $\tilde{\boldsymbol{\theta}}_n$ by

$$\tilde{\sigma}_n^2 = \frac{\tilde{S}_n(\tilde{\boldsymbol{\theta}}_n)}{n}, \tag{4.4}$$

where

$$\tilde{S}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \frac{(X_t - \tilde{X}_t(\boldsymbol{\theta}))^2}{\tilde{r}_t(\boldsymbol{\theta})}.$$

Substitution of (4.4) into (4.3) then yields that $\tilde{\boldsymbol{\theta}}_n$ is a maximizer of the so–called reduced log–likelihood

$$\tilde{Q}_n(\boldsymbol{\theta}) = -\log(n^{-1}\tilde{S}_n(\boldsymbol{\theta})) - n^{-1} \sum_{t=1}^{n} \log \tilde{r}_t(\boldsymbol{\theta}). \tag{4.5}$$

Hence the latter trick eliminates the scale parameter σ^2 from the quasi log–likelihood and thereby reduces the dimension of the parameter space, in which one has to find a maximum, by 1. By the same arguments, σ^2 can be removed from the approximate conditional Gaussian log–likelihood $\hat{L}_n$, which will be defined in equation (4.7) below. □

Conditional Gaussian Likelihood

In what follows, we explain the derivation of an estimator which is asymptotically equivalent to $\tilde{\boldsymbol{\theta}}_n$. The present exposition will be helpful when it comes to derive an *approximate* conditional Gaussian likelihood in the GARCH model.

Instead of using the exact Gaussian log–likelihood (4.1), we concentrate on finding an approximation to a conditional Gaussian likelihood. Under parameter $\boldsymbol{\theta}$ and innovations variance σ^2 we have that the vector $(X_1, \ldots, X_n)^T$ given $X_0, \ldots, X_{-p+1}, Z_0, \ldots, Z_{-q+1}$ is conditionally multivariate normally distributed with conditional Gaussian log–likelihood

$$\check{L}_n(\boldsymbol{\theta}, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(X_t - \check{X}_t(\boldsymbol{\theta}))^2. \tag{4.6}$$

Here

$$\check{X}_t(\boldsymbol{\theta}) = \sum_{i=1}^{p}\varphi_i X_{t-i}(\boldsymbol{\theta}) + \sum_{j=1}^{q}\vartheta_j \check{Z}_{t-j}(\boldsymbol{\theta}),$$

where the random variables

$$\check{Z}_t(\boldsymbol{\theta}) = \begin{cases} Z_t, & t \le 0, \\ X_t - \sum_{i=1}^{p}\varphi_i X_{t-i} - \sum_{j=1}^{q}\vartheta_j \check{Z}_{t-j}(\boldsymbol{\theta}), & t > 0, \end{cases}$$

are the innovations at time t under the parameter hypothesis $\boldsymbol{\theta}$. In particular, note that $\check{Z}_t(\boldsymbol{\theta}_0) = Z_t = X_t - \check{X}_t(\boldsymbol{\theta}_0)$ when $t > 0$.

Since in practice $Z_0, \ldots, Z_{-q+1}$ are unobservable and $X_0, \ldots, X_{-p+1}$ are not available, the quantities $\check{X}_t(\boldsymbol{\theta})$ and $\check{Z}_t(\boldsymbol{\theta})$ have to be approximated in some way. Since $\mathbb{E}X_t = \mathbb{E}Z_t = 0$, a natural choice is to impute $X_t = Z_t = 0$ for all $t \le 0$, which then leads to

$$\hat{Z}_t(\boldsymbol{\theta}) = X_t - \sum_{i=1}^{\min(p,t-1)}\varphi_i X_{t-i} - \sum_{j=1}^{\min(q,t-1)}\vartheta_j \hat{Z}_{t-j}(\boldsymbol{\theta}), \qquad t > 0,$$

as an approximation of $\check{Z}_t(\boldsymbol{\theta})$, and

$$\hat{X}_t(\boldsymbol{\theta}) = \sum_{i=1}^{\min(p,t-1)}\varphi_i X_{t-i} + \sum_{j=1}^{\min(q,t-1)}\vartheta_j \hat{Z}_{t-j}(\boldsymbol{\theta}), \qquad t > 0,$$

as an approximation of $\check{X}_t(\boldsymbol{\theta})$. We now replace $\check{X}_t(\boldsymbol{\theta})$ in (4.6) by $\hat{X}_t(\boldsymbol{\theta})$ and so obtain an approximation to the conditional Gaussian log–likelihood $\check{L}_n$, namely

$$\hat{L}_n(\boldsymbol{\theta}, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(X_t - \hat{X}_t(\boldsymbol{\theta}))^2\,. \tag{4.7}$$

Denote by $\binom{\hat{\boldsymbol{\theta}}_n}{\hat{\sigma}_n^2}$ the maximizer of $\hat{L}_n$. As we have alluded to already, the two estimators $\hat{\boldsymbol{\theta}}_n$ and $\tilde{\boldsymbol{\theta}}_n$ are asymptotically equivalent, i.e., $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ and $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ have the same weak limit as $n \to \infty$.

Corollary 4.1.4. *Under the assumptions of Theorem 4.1.1, one has that* $\hat{\sigma}_n^2 \xrightarrow{\text{a.s.}} (\sigma^\circ)^2$ *and the estimators* $\hat{\boldsymbol{\theta}}_n$ *and* $\tilde{\boldsymbol{\theta}}_n$ *are asymptotically equivalent since*

$$\sqrt{n}\,|\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n| \xrightarrow{\text{a.s.}} 0, \qquad n \to \infty.$$

Sketch of Proof. We are not aware of a proof to this corollary in the literature. Since a complete proof is beyond the scope of this monograph, we restrict ourselves to explaining the main ideas. One can show (see Problem 5.15 in Brockwell and Davis [29]) that

$$\hat{Z}_t(\boldsymbol{\theta}) = \sum_{j=0}^{t-1} \pi_j(\boldsymbol{\theta}) X_{t-j}, \qquad t > 0,$$

where $\pi_j(\boldsymbol{\theta})$ is defined through

$$\sum_{j=0}^{\infty} \pi_j(\boldsymbol{\theta}) z^j = \frac{\varphi(z)}{\vartheta(z)} = \frac{1 - \varphi_1 z - \cdots - \varphi_p z^p}{1 + \vartheta_1 z + \cdots + \vartheta_q z^q}, \qquad |z| \le 1.$$

This suggests the following definitions:

$$Z_t(\boldsymbol{\theta}) = \sum_{j=0}^{\infty} \pi_j(\boldsymbol{\theta}) X_{t-j}, \qquad t \in \mathbb{Z},$$

and

$$X_t(\boldsymbol{\theta}) = \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{j=1}^{q} \vartheta_j Z_{t-q}(\boldsymbol{\theta}), \qquad t \in \mathbb{Z}.$$

Recalling equation (3.4) reveals that $X_t(\boldsymbol{\theta})$ coincides with the best linear predictor of X_t based on the infinite past, i.e.,

$$X_t(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta},\sigma^2}(X_t \,|\, X_{t-1}, X_{t-2}, \ldots), \qquad t \in \mathbb{Z}.$$

These observations are useful for understanding the asymptotic equivalence of $\tilde{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n$, since they imply that

$$|\hat{Z}_t(\boldsymbol{\theta}) - Z_t(\boldsymbol{\theta})| = \left| \sum_{j=t}^{\infty} \pi_j(\boldsymbol{\theta}) X_{t-j} \right| \le \sum_{j=t}^{\infty} |\pi_j(\boldsymbol{\theta})|\, |X_{t-j}|, \qquad t > 0.$$

Thus there is a constant $c > 0$ such that

$$|\hat{X}_t(\boldsymbol{\theta}) - X_t(\boldsymbol{\theta})| \le c \sum_{j=t-q}^{\infty} |\pi_j(\boldsymbol{\theta})|\, |X_{t-j}|, \qquad t > \max(p,q). \tag{4.8}$$

An application of the Cauchy inequalities (see Rudin [120]) shows that $|\pi_j(\boldsymbol{\theta})|$ decays to zero exponentially fast as $j \to \infty$. Since $\mathbb{E}|X_0| < \infty$ implies $\mathbb{E}(\log^+ |X_0|) < \infty$ (Lemma 2.5.3), an application of Proposition 2.5.1 to the right–hand side of inequality (4.8) demonstrates that the error $|\hat{X}_t(\boldsymbol{\theta}) - X_t(\boldsymbol{\theta})|$ decays to zero exponentially fast with probability 1, i.e.,

$$|\hat{X}_t(\boldsymbol{\theta}) - X_t(\boldsymbol{\theta})| \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty.$$

From this one concludes that $\hat{L}_n(\boldsymbol{\theta}, \sigma^2) = L_n(\boldsymbol{\theta}, \sigma^2) + R_n$ with $n^{-1} R_n \xrightarrow{\text{a.s.}} 0$, where

$$L_n(\boldsymbol{\theta}, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{n} (X_t - X_t(\boldsymbol{\theta}))^2 .$$

In fact one can even show that, uniformly on a suitable compact set containing the true parameter vector, $n^{-1}|\hat{L}_n - L_n| \xrightarrow{\text{a.s.}} 0$ and $n^{-1/2}|\hat{L}'_n - L'_n| \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$ ($'$ denotes the derivative with respect to $(\boldsymbol{\theta}, \sigma^2)^T$). The same limit results hold true for the differences $|\tilde{L}_n - L_n|$. Hence

$$n^{-1}|\hat{L}_n - \tilde{L}_n| \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad n^{-1/2}|\hat{L}'_n - \tilde{L}'_n| \xrightarrow{\text{a.s.}} 0, \qquad n \to \infty. \tag{4.9}$$

Then an evident adaptation of the proofs of Theorem 5.3.1 and Lemma 5.6.5 shows that the limit relation (4.9) implies $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n) \xrightarrow{\text{a.s.}} 0$. □

4.1.2 Least–squares Estimation

The least–squares estimator is a minimizer $\tilde{\boldsymbol{\theta}}_n^{\text{LS}}$ of the quantity

$$\tilde{S}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \frac{(X_t - \tilde{X}_t(\boldsymbol{\theta}))^2}{\tilde{r}_t(\boldsymbol{\theta})},$$

which already appeared in the reduced log–likelihood (4.5). In the literature sometimes also the minimizer $\tilde{\tilde{\boldsymbol{\theta}}}_n^{\text{LS}}$ of

$$\tilde{\tilde{S}}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} (X_t - \tilde{X}_t(\boldsymbol{\theta}))^2 .$$

is referred to as a least–squares estimator. It can be shown (see the proof of Theorem 10.8.2 in Brockwell and Davis [29]) that under the assumptions of Theorem 4.1.1 one has $\tilde{\boldsymbol{\theta}}_n^{\text{LS}} - \tilde{\boldsymbol{\theta}}_n = o_{\mathbb{P}}(n^{-1/2})$. This implies that $\tilde{\boldsymbol{\theta}}_n^{\text{LS}}$ and $\tilde{\boldsymbol{\theta}}_n$ are asymptotically equivalent. Using the techniques exploited for the proofs of Theorem 10.8.2 and Propositions 10.8.3 and 10.8.4 in Brockwell and Davis [29] it is also possible to establish $\tilde{\tilde{\boldsymbol{\theta}}}_n^{\text{LS}} - \tilde{\boldsymbol{\theta}}_n = o_{\mathbb{P}}(n^{-1/2})$. As a remark we mention that the asymptotic properties of $\tilde{\tilde{\boldsymbol{\theta}}}_n^{\text{LS}}$ were also studied by Klimko and Nelson [75] and Tjøstheim [127] in the context of general (nonlinear) time series models.

4.1.3 Whittle Estimation

Whittle [133] devised an approximation to the Gaussian log–likelihood $\tilde{L}_n$ based on the so–called periodogram. So far we have worked in the time domain, whereas the periodogram is an estimator used if one concentrates on the frequency domain, or equivalently, if one conducts spectral analysis of time series. In what follows, we briefly summarize the main ideas and facts about spectral time series analysis. We refer to Priestley [114] or to Chapter 4 of Brockwell and Davis [29] for a treatment in full mathematical generality. A further reference for Whittle estimation is Dzhaparidze [44].

Spectral Distribution Function and Spectral Density

At the origin of spectral theory lies the spectral representation theorem, which states that any stationary stochastic process (X_t) with mean zero and finite variance admits an almost sure stochastic integral representation

$$X_t = \int_{(-\pi,\pi]} e^{i\lambda t} \mathrm{d}U(\lambda), \qquad t \in \mathbb{Z}, \tag{4.10}$$

with respect to a mean–zero complex valued process $(U(\lambda))_{\lambda\in[-\pi,\pi]}$ with uncorrelated increments and such that

$$\mathbb{E}|U(\lambda_2) - U(\lambda_1)|^2 = F(\lambda_2) - F(\lambda_1), \qquad -\pi \le \lambda_1 \le \lambda_2 \le \pi,$$

where F is a nondecreasing right–continuous bounded function on $[-\pi,\pi]$ and standardized to $F(-\pi) = 0$. The function $F(\lambda)$, $\lambda \in [-\pi,\pi]$, is called the spectral distribution function of the autocovariance function $\gamma_X(\cdot)$. One can show that the relationship

$$\gamma_X(h) = \int_{(-\pi,\pi]} e^{ih\lambda} \mathrm{d}F(\lambda), \qquad h \in \mathbb{Z},$$

is valid and that there is a one–to–one relationship between autocovariance functions and spectral distribution functions; for these reasons, time domain results have equivalent frequency domain counterparts. If F is absolutely continuous with respect to Lebesgue measure, the derivative $f(\lambda) = F'(\lambda)$ is called spectral density. For example, if the autocovariances are absolutely summable, then F is absolutely continuous. In that case the spectral density has Fourier series representation

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \gamma_X(h) e^{-ih\lambda} \\ &= \frac{1}{2\pi} \left(\gamma_X(0) + 2\sum_{h=1}^{\infty} \gamma_X(h)\cos(h\lambda) \right), \qquad \lambda \in (-\pi,\pi]. \end{aligned} \tag{4.11}$$

Roughly speaking, the spectral representation theorem (4.10) says that X_t is a "weighted superposition" of trigonometric functions $\exp(i\lambda t)$, $\lambda \in (-\pi, \pi]$, with "zero–mean uncorrelated random weights, which have variances proportional to $f(\lambda)$" (provided F is absolutely continuous).

Our aim is to compute the spectral density of a causal ARMA process (X_t) with $\mathrm{Var}(Z_0) = \sigma^2 < \infty$ and parameter vector $\boldsymbol{\theta} = (\varphi_1, \ldots, \varphi_p, \vartheta_1, \ldots, \vartheta_q)^T$. To this end, one exploits the fact that the application of a time–invariant linear filter $\eta(B) = \sum_{j=0}^{\infty} \eta_j B^j$ to a stationary process (Y_t) with spectral density $f_Y(\lambda)$ corresponds to a multiplication of $f_Y(\lambda)$ with the associated power transfer function

$$|\eta(e^{-i\lambda})|^2 = \Big| \sum_{j=0}^{\infty} \eta_j e^{-ij\lambda} \Big|^2, \qquad \lambda \in (-\pi, \pi].$$

Indeed, if $\sum_{j=0}^{\infty} |\eta_j| < \infty$, then by Theorem 4.4.1 in Brockwell and Davis [29] the process

$$V_t = \eta(B) Y_t = \sum_{j=0}^{\infty} \eta_j Y_{t-j}, \qquad t \in \mathbb{Z},$$

has spectral density

$$f_V(\lambda) = |\eta(e^{-i\lambda})|^2 \, f_Y(\lambda), \qquad \lambda \in (-\pi, \pi]. \tag{4.12}$$

Now recall from Section 3.2 that a causal ARMA process (X_t) with parameter vector $\boldsymbol{\theta} = (\varphi_1, \ldots, \varphi_p, \vartheta_1, \ldots, \vartheta_q)^T$ has the representation

$$X_t = \varphi(B)^{-1} \vartheta(B) Z_t = \psi(B) Z_t, \qquad t \in \mathbb{Z},$$

where

$$\psi(z) = \frac{\vartheta(z)}{\varphi(z)} = \frac{1 + \vartheta_1 z + \cdots + \vartheta_q z^q}{1 - \varphi_1 z - \cdots - \varphi_p z^p} = \sum_{j=0}^{\infty} \psi_j z^j, \qquad |z| \le 1.$$

Then, since (Z_t) has spectral density $f_Z(\lambda) = \sigma^2/(2\pi)$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$, the spectral density of (X_t) equals

$$f_X(\lambda; \boldsymbol{\theta}) = \frac{\sigma^2}{2\pi} \frac{|\vartheta(e^{-i\lambda})|^2}{|\varphi(e^{-i\lambda})|^2} = \frac{\sigma^2}{2\pi} \, g(\lambda; \boldsymbol{\theta}), \qquad \lambda \in (-\pi, \pi], \tag{4.13}$$

by virtue of relation (4.12); notice that the expression $g(\lambda; \boldsymbol{\theta})$ has already appeared in Theorem 4.1.1, where it determined the asymptotic covariance matrix $[\mathbf{W}(\boldsymbol{\theta}_0)]^{-1}$.

The following observation turns out to be essential for the definition of the Whittle estimator. Suppose $\boldsymbol{\theta}_0 \in C$; then

$$\int_{-\pi}^{\pi} \frac{g(\lambda; \boldsymbol{\theta}_0)}{g(\lambda; \boldsymbol{\theta})} \, \mathrm{d}\lambda > 2\pi, \qquad \text{for all } \boldsymbol{\theta} \in \bar{C},\ \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \tag{4.14}$$

or in other words, the function $\boldsymbol{\theta} \mapsto \int_{-\pi}^{\pi} g(\lambda; \boldsymbol{\theta}_0)/g(\lambda; \boldsymbol{\theta}) \, \mathrm{d}\lambda$ on $\bar{C}$ is uniquely minimized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (Proposition 10.8.1 in Brockwell and Davis [29]).

The Periodogram

The periodogram of the (mean-corrected) sample $X_1, \ldots, X_n$ is defined as

$$I_{n,X}(\lambda) = \frac{1}{n} \left| \sum_{t=1}^{n} (X_t - \overline{X})\, e^{-i\lambda t} \right|^2, \qquad \lambda \in (-\pi, \pi], \tag{4.15}$$

where $\overline{X}$ denotes the sample mean. It is straightforward to verify that alternatively

$$I_{n,X}(\lambda) = \sum_{|h|<n} \gamma_{n,X}(h) e^{-ih\lambda},$$

which shows that the periodogram is (up to the multiplicative constant 2π) nothing but a moment estimator for the spectral density of the underlying time series (X_t), i.e., it is a sample analogue of the spectral density (4.11). Since this monograph does not primarily deal with the asymptotic properties of the periodogram, we refer to Chapter 10.3 of Brockwell and Davis [29] for a comprehensive discussion. We merely mention that the periodogram is often evaluated at the Fourier frequencies

$$\lambda_j = \frac{2\pi j}{n}, \qquad j = -[(n-1)/2], \ldots, [n/2], \tag{4.16}$$

since then numerical computations can be made via the fast Fourier transform; see Chapter 10.7 in Brockwell and Davis [29].

Definition of the Whittle Estimator

We have now collected the necessary items which make us understand the definition of the Whittle estimator. The property (4.14) suggests that

$$\boldsymbol{\theta} \mapsto \int_{-\pi}^{\pi} \frac{g(\lambda; \boldsymbol{\theta}_0)}{g(\lambda; \boldsymbol{\theta})}\, \mathrm{d}\lambda \tag{4.17}$$

is a suitable objective function to be minimized. A naive estimation procedure for $\boldsymbol{\theta}_0$ is therefore given as follows:

- Replace the (unknown) spectral density $g(\lambda; \boldsymbol{\theta}_0)$ under the integral (4.17) by its sample version, the periodogram.
- Replace the integral $\int_{-\pi}^{\pi} \cdot$ in (4.17) by a Riemann sum evaluated at the Fourier frequencies (4.16):

$$\bar{\sigma}^2_{n,X}(\boldsymbol{\theta}) = \frac{1}{n} \sum_j \frac{I_{n,X}(\lambda_j)}{g(\lambda_j; \boldsymbol{\theta})}.$$

  (The summation is taken over all Fourier frequencies.)
- Minimize $\bar{\sigma}^2_{n,X}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta} \in \bar{C}$.

This leads to the Whittle estimator given by

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{W}} = \operatorname*{argmin}_{\boldsymbol{\theta} \in \bar{C}} \bar{\sigma}^2_{n,X}(\boldsymbol{\theta}) \,. \tag{4.18}$$

We mention that Whittle's [133] derivation of $\hat{\boldsymbol{\theta}}_n^{\mathrm{W}}$ was rather motivated through an approximation to the Gaussian log–likelihood $\tilde{L}_n$, whereas our way of explaining $\hat{\boldsymbol{\theta}}_n^{\mathrm{W}}$ is mainly influenced by Hannan's [64] proof of consistency. Since $|\hat{\boldsymbol{\theta}}_n^{\mathrm{W}} - \tilde{\boldsymbol{\theta}}_n| = o_{\mathbb{P}}(n^{-1/2})$ as $n \to \infty$ (see the proof of Theorem 10.8.2 in Brockwell and Davis [29]), the estimators $\hat{\boldsymbol{\theta}}_n^{\mathrm{W}}$ and $\tilde{\boldsymbol{\theta}}_n$ are asymptotically equivalent.

Theorem 4.1.5 (Hannan [64]). *Let (X_t) be a causal invertible* ARMA(p,q) *process with true parameters $\boldsymbol{\theta}_0 \in C$ and $0 < \mathrm{Var}(Z_0) = (\sigma^\circ)^2 < \infty$. Then*

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{W}} \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0, \qquad \bar{\sigma}^2_{n,X}(\hat{\boldsymbol{\theta}}_n^{\mathrm{W}}) \xrightarrow{\text{a.s.}} (\sigma^\circ)^2, \qquad n \to \infty,$$

and

$$\sqrt{n}\,(\hat{\boldsymbol{\theta}}_n^{\mathrm{W}} - \boldsymbol{\theta}_0) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, [\mathbf{W}(\boldsymbol{\theta}_0)]^{-1}), \qquad n \to \infty,$$

where $\mathbf{W}(\boldsymbol{\theta}_0)$ is as in Theorem 4.1.1.

4.2 Estimation for GARCH Processes

We discuss the (Gaussian) quasi maximum likelihood estimator (QMLE) in the GARCH(p,q) model and the Whittle estimator applied to the squares of a GARCH(p,q) process (X_t). In contrast to ARMA processes discussed in the previous sections, the two procedures do *not* lead to asymptotically equivalent estimators for the GARCH(p,q) parameters. This phenomenon stems from the fact that the Whittle estimator is applied to transformed data, i.e, the squares (X_t^2), which formally constitute an ARMA$(\max(p,q),q)$ process with a certain white noise sequence. For this reason, a direct comparison of the QMLE and Whittle estimator is rather inappropriate, and the asymptotic inequivalence of the two estimators is not totally unexpected.

As a matter of fact, it is beneficial to use the QMLE because it is much less sensitive with respect to a heavy tailed unconditional distribution of $|X_0|$ than Whittle. For consistency, the Whittle estimator requires a finite $4th$ unconditional moment, i.e., $\mathbb{E}X_0^4 < \infty$, whereas the QMLE is consistent for *all* stationary GARCH(p,q) processes, provided the distribution of Z_0 is not concentrated in two points. If also $\mathbb{E}Z_0^4 < \infty$, then the QMLE is even asymptotically normal. In contrast, the Whittle estimator in GARCH(p,q) is only asymptotically normal if $\mathbb{E}X_0^8 < \infty$ and has slower than $\sqrt{n}$–rate of convergence in the case of a tail index $\kappa \in (4,8)$. As an illustration of these statements, let us consider parameter estimation in an IGARCH$(1,1)$ model, i.e., a GARCH$(1,1)$ model where $\alpha_1 + \beta_1 = 1$; suppose the innovations (Z_t)

are standard Gaussian. We have seen on p. 60 that the distribution of $|X_0|$ has tail index $\kappa = 2$ and $\mathbb{E}X_0^2 = \infty$. Thus the QMLE applied to IGARCH$(1,1)$ is asymptotically normal, whereas Whittle is inconsistent.

Instead of Whittle estimation one might also want to apply the least–squares estimator to (X_t^2). Since the arguments of Brockwell and Davis [29] for the asymptotic equivalence of Whittle and least–squares estimator remain valid in the GARCH case, the least–squares estimator has the same limit behavior as the Whittle estimator. For this reason we may without loss of generality focus on an analysis of the Whittle estimator.

4.2.1 Quasi Maximum Likelihood Estimation

The quasi maximum likelihood estimator (QMLE) is the most common procedure for estimating GARCH parameters. Most statistical computer packages or specialized software for financial time series analysis contain the QMLE as a built–in function. But despite its popularity, a mathematical proof for the consistency and asymptotic normality had not been provided for a long time. Recently, Berkes et al. [8] established consistency and asymptotic normality of the QMLE in GARCH(p,q) under weak assumptions on the parameters and the distribution of the underlying noise sequence (Z_t). Thereby they generalized work by Lee and Hansen [84] and Lumsdaine [90] on GARCH$(1,1)$. Straumann and Mikosch [125] provided a unifying theory of the QMLE in conditionally heteroscedastic time series models and applied it to AGARCH(p,q); the latter theory is presented in Chapter 5 of this monograph.

In what follows, we show how one can derive an approximation to the conditional Gaussian likelihood of a stationary GARCH(p,q) process (X_t), i.e.,

$$X_t = \sigma_t Z_t, \qquad t \in \mathbb{Z},$$

where (Z_t) is a sequence of iid random variables with $\mathbb{E}Z_0 = 0$ and $\mathrm{Var}(Z_0) = 1$, and

$$\sigma_t^2 = \alpha_0^\circ + \sum_{i=1}^{p} \alpha_i^\circ X_{t-i}^2 + \sum_{j=1}^{q} \beta_j^\circ \sigma_{t-j}^2, \qquad t \in \mathbb{Z}. \tag{4.19}$$

Denote the true parameter vector by $\boldsymbol{\theta}_0 = (\alpha_0^\circ, \ldots, \alpha_p^\circ, \beta_1^\circ, \ldots, \beta_q^\circ)^T$. In contrast to ARMA, even under the synthetic assumption that the Z_t's are iid $\sim \mathcal{N}(0,1)$, there is no explicit expression of the probability density of a GARCH(p,q) vector $(X_1, \ldots, X_n)^T$ since the distribution of $(\sigma_1, \ldots, \sigma_n)^T$ is not known. To overcome this difficulty, one can consider an approximate *conditional* Gaussian log–likelihood instead, similarly to the approach which has lead to $\hat{\boldsymbol{\theta}}_n$ in Section 4.1.1.

Let $\boldsymbol{\theta} = (\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^T$ denote the presumed parameter and make use of the decomposition

$$\log f_{\boldsymbol{\theta}}(X_1,\ldots,X_n \mid X_0,\ldots,X_{-p+1},\sigma_0^2,\ldots,\sigma_{-q+1}^2) = \sum_{t=1}^{n} \log f_{\boldsymbol{\theta}}(X_t \mid X_{t-1},\ldots,X_{-p+1},\sigma_0^2,\ldots,\sigma_{-q+1}^2)$$

in order to observe that X_t given $X_{t-1},\ldots,X_{-p+1}$ and $\sigma_0^2,\ldots,\sigma_{-q+1}^2$ is conditionally Gaussian distributed with mean zero and variance $\check{h}_t(\boldsymbol{\theta})$ recursively given by

$$\check{h}_t(\boldsymbol{\theta}) = \begin{cases} \sigma_t^2, & t \le 0, \\ \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \check{h}_{t-j}(\boldsymbol{\theta}), & t > 0. \end{cases}$$

This implies that the conditional Gaussian log–likelihood has the form

$$\log f_{\boldsymbol{\theta}}(X_1,\ldots,X_n|X_0,\ldots,X_{-p+1},\sigma_0^2,\ldots,\sigma_{-q+1}^2) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left(\frac{X_t^2}{\check{h}_t(\boldsymbol{\theta})} + \log \check{h}_t(\boldsymbol{\theta})\right). \quad (4.20)$$

Since $X_0,\ldots,X_{-p+1}$ are unavailable and the squared volatilities $\sigma_0^2,\ldots,\sigma_{-q+1}^2$ unobservable, the conditional Gaussian log–likelihood (4.20) cannot be numerically evaluated without a certain initialization for $\sigma_0^2,\ldots,\sigma_{-p+1}^2$ and $X_0,\ldots,X_{-q+1}$. It will follow in Chapter 5 that the initial values are asymptotically irrelevant so that we set the X_t's equal to zero and $\hat{h}_t(\boldsymbol{\theta}) = \alpha_0/(1-\beta_1-\cdots-\beta_q)$ for $t \le 0$; formula (4.26) below together with (4.28) is a possible justification of our particular choice for $\hat{h}_t(\boldsymbol{\theta})$, $t \le 0$. We arrive at

$$\hat{h}_t(\boldsymbol{\theta}) = \begin{cases} \alpha_0/(1-\beta_1-\cdots-\beta_q), & t \le 0, \\ \alpha_0 + \sum_{i=1}^{\min(p,t-1)} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \hat{h}_{t-j}(\boldsymbol{\theta}), & t > 0. \end{cases} \quad (4.21)$$

The function $(\hat{h}_t(\boldsymbol{\theta}))^{1/2}$ can be understood as an estimate of the volatility at time t and under parameter hypothesis $\boldsymbol{\theta}$. The results of Chapter 5 imply that, uniformly on the compact set K defined in (4.22) below, $|\hat{h}_t - \check{h}_t| \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$. This suggests that by replacing $\check{h}_t(\boldsymbol{\theta})$ by $\hat{h}_t(\boldsymbol{\theta})$ in (4.20) we obtain a good approximation to the conditional Gaussian log–likelihood. Since the constant $-n\log(2\pi)/2$ does not matter for the optimization, we define the QMLE $\hat{\boldsymbol{\theta}}_n$ as a maximizer of the function

$$\hat{L}_n(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{n}\left(\frac{X_t^2}{\hat{h}_t(\boldsymbol{\theta})} + \log \hat{h}_t(\boldsymbol{\theta})\right)$$

with respect to $\boldsymbol{\theta} \in K$, where the compact set K fulfills

$$K \subset (0,\infty) \times [0,\infty)^p \times B \quad (4.22)$$

with

$$B = \{(\beta_1, \dots, \beta_q)^T \in [0,1)^q \mid \sum_{j=1}^{q} \beta_j < 1\}.$$

We now list several conditions, which will be needed for the consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_n$:

Q.1 The distribution of Z_0 is not concentrated in two points.

Q.2 $(\alpha_p^\circ, \beta_q^\circ) \neq (0,0)$, $\alpha_0^\circ > 0$ and there is $i > 0$ with $\alpha_i^\circ > 0$. The polynomials $\alpha^\circ(z) = \alpha_1^\circ z + \cdots + \alpha_p^\circ z^p$ and $\beta^\circ(z) = 1 - \beta_1^\circ z - \cdots - \beta_q^\circ z^q$ do not have any common zeros.

Q.3 The true parameter $\boldsymbol{\theta}_0$ lies in the interior of K.

Q.4 There is $\mu > 0$ such that $\mathbb{P}(|Z_0| \leq t) = o(t^\mu)$ as $t \downarrow 0$.

We are now ready to quote Theorems 4.1 and 4.2 of Berkes et al. [8]. We present the results in a slightly more general form.

Theorem 4.2.1 (Berkes et al. [8]). *Let (X_t) be a stationary* GARCH(p,q) *process with true parameter vector $\boldsymbol{\theta}_0 \in K$. Suppose the conditions* Q.1 *and* Q.2 *hold. Then the* QMLE $\hat{\boldsymbol{\theta}}_n$ *is strongly consistent, i.e.,*

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0, \qquad n \to \infty.$$

If in addition $\mathbb{E}Z_0^4 < \infty$ and Q.3 *and* Q.4 *hold true, the* QMLE $\hat{\boldsymbol{\theta}}_n$ *is also asymptotically normal, i.e.,*

$$\sqrt{n}\,(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, \mathbf{F}_0^{-1}\mathbf{G}_0\mathbf{F}_0^{-1}),$$

where the $(p+q+1) \times (p+q+1)$ matrices $\mathbf{G}_0$ and $\mathbf{F}_0$ are given by

$$\mathbf{F}_0 = -\frac{2}{\mathbb{E}(Z_0^4 - 1)}\mathbf{G}_0,$$

$$\mathbf{G}_0 = \frac{\mathbb{E}(Z_0^4 - 1)}{4}\,\mathbb{E}\left(\frac{1}{\sigma_0^4}\left(\frac{\partial h_0(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)^T \frac{\partial h_0(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right).$$

The $\mathbb{C}(K)$–valued stochastic process (h_t) is the unique stationary solution of the difference equation

$$h_t(\boldsymbol{\theta}) = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}(\boldsymbol{\theta}), \qquad t \in \mathbb{Z}. \tag{4.23}$$

Remark 4.2.2. Berkes et al. [8] even require $\mathbb{E}|Z_0|^{4+\epsilon} < \infty$ for some $\epsilon > 0$. It seems however that $\mathbb{E}Z_0^4 < \infty$ is sufficient since the proof of their Theorem 4.2 goes through under the weaker condition $\mathbb{E}Z_0^4 < \infty$. □

A Discussion of the Conditions Q.1 – Q.4

- The conditions Q.1 and Q.4 are rather mild technical assumptions. Note that the requirement $\mathbb{P}(|Z_0| \leq t) = o(t^\mu)$ as $t \to 0$ prevents the distribution of $|Z_0|$ from having too much mass around zero. It is e.g. fulfilled if Z_0 has a bounded Lebesgue density in some neighborhood of zero.

- Observe that the set K given in (4.22) may contain parameters $\boldsymbol{\theta}$ that do *not* belong to a stationary GARCH(p,q) model. Crucial in the definition of K is however the restriction $\beta_1 + \cdots + \beta_q \leq \bar{\beta} < 1$ and $\beta_j \geq 0$ for all $\boldsymbol{\theta} \in K$; the latter restriction is necessary for the existence and uniqueness of a stationary solution to (4.23) or equivalently the validity of representation (4.26) below.

- Condition Q.2 is needed to ensure the identifiability of the GARCH(p,q) parameters, i.e., that there is no $\tilde{\boldsymbol{\theta}}_0 = (\tilde{\alpha}_0^\circ, \tilde{\alpha}_1^\circ, \ldots, \tilde{\alpha}_p^\circ, \tilde{\beta}_1^\circ, \ldots, \tilde{\beta}_q^\circ)^T \neq \boldsymbol{\theta}_0$ such that
$$\sigma_t^2 = \tilde{\alpha}_0^\circ + \sum_{i=1}^p \tilde{\alpha}_i^\circ X_{t-i}^2 + \sum_{j=1}^q \tilde{\beta}_j^\circ \sigma_{t-j}^2 \qquad \text{a.s.}$$
The exclusion of $\alpha_0^\circ = 0$ is evident since $\alpha_0^\circ = 0$ implies $\sigma_t^2 \equiv 0$. If $\alpha_i^\circ = 0$ for all $i \in [1,p]$, we would have $\sigma_t^2 = \alpha_0^\circ/(1 - \sum_{j=1}^q \beta_j^\circ)$, which implies that parameters lying on the curve $\alpha_0/(1 - \sum_{j=1}^q \beta_j^\circ) = \text{const}$ lead to identical models. Observe that the difference equation (4.19) for the squared volatility of a GARCH(p,q) process can be written as
$$\beta^\circ(B)\sigma_t^2 = \alpha^\circ(B)\left(\frac{\alpha_0}{\alpha^\circ(1)} + X_t^2\right) \qquad (4.24)$$
in backshift operator notation, where $\alpha^\circ(z)$ and $\beta^\circ(z)$ are the polynomials defined in condition Q.2. Analogously to ARMA models, the occurrence of common zeros in the polynomials $\alpha^\circ(z)$ and $\beta^\circ(z)$ would imply that nontrivial common factors in $\alpha^\circ(z)$ and $\beta^\circ(z)$ cancel out, which leads to an identifiability problem. If $(\alpha_p^\circ, \beta_q^\circ) = (0,0)$, then multiplication of both sides of (4.24) with an appropriate linear factor $(aB + c)$ shows that there is $\tilde{\boldsymbol{\theta}}_0 \neq \boldsymbol{\theta}_0$ belonging to the identical GARCH(p,q) model.

- Condition Q.3 rules out that $\boldsymbol{\theta}_0$ lies on the boundary of the set of GARCH(p,q) parameters belonging to a stationary model. For boundary points asymptotic normality of the QMLE is of course impossible: if e.g. the QMLE of GARCH(1,1) is applied to ARCH(1) data, then $\beta_1^\circ = 0$ and $\hat{\beta}_1^{(n)} \geq 0$ (label $\hat{\boldsymbol{\theta}}_n = (\hat{\alpha}_0^{(n)}, \hat{\alpha}_1^{(n)}, \hat{\beta}_1^{(n)})^T$) so that the coordinate $\hat{\beta}_1^{(n)}$ cannot be asymptotically normal. Q.3 also implies that the order (p,q) of the GARCH model must be *perfectly* specified for asymptotic normality. We mention that Berkes et al. [8] work under the assumption Q.3 also for consistency. This restriction is obsolete if the techniques of Jeantheau [70], which go back to Pfanzagl [112], are employed for the proof of consistency.

Moreover, Berkes et al. [8] assume $\mathbb{E}|Z_0|^{2+\delta}$ for a $\delta > 0$. This is again not necessary if one uses the consistency proof of Pfanzagl [112]; cf. the proof of Theorem 5.3.1 in this monograph.

Some Remarks Concerning the Proof of Theorem 4.2.1

Here we merely sketch the main steps of proof of Theorem 4.2.1. In Chapter 5 we develop a general theory for quasi maximum likelihood inference in conditionally heteroscedastic time series models which may be applied to GARCH(p, q) for obtaining Theorem 4.2.1.

The proof essentially consists of a careful analysis of the asymptotic behavior of $(\hat{L}_n)_{n\geq 1}$ and the corresponding sequences of first and second derivatives. As a first step, one seeks to find a stationary approximation to the sequence $(\log \hat{h}_t + X_t^2/\hat{h}_t)_{t\in\mathbb{N}}$ with an asymptotically negligible error. It turns out that one has to replace $\hat{h}_t$ in $\hat{L}_n$ by the unique stationary solution of the difference equation (4.23):

$$L_n(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{n}\left(\frac{X_t^2}{h_t(\boldsymbol{\theta})} + \log h_t(\boldsymbol{\theta})\right). \tag{4.25}$$

Then one analyzes the maximizer of L_n, which is asymptotically equivalent to $\hat{\boldsymbol{\theta}}_n$. In what follows, we give some additional details.

As regards the difference equation (4.23), note that in backshift operator notation it becomes

$$\beta_{\boldsymbol{\theta}}(B)h_t(\boldsymbol{\theta}) = \alpha_{\boldsymbol{\theta}}(B)\left(\frac{\alpha_0}{\alpha_{\boldsymbol{\theta}}(1)} + X_t^2\right),$$

where $\alpha_{\boldsymbol{\theta}}(z) = \alpha_1 z + \cdots + \alpha_p z^p$ and $\beta_{\boldsymbol{\theta}}(z) = 1 - \beta_1 z - \cdots - \beta_q z^q$. Observe that the conditions $\beta_1 + \cdots + \beta_q \leq \bar{\beta} < 1$ and $\beta_j \geq 0$ imply that the polynomial $\beta_{\boldsymbol{\theta}}(z)$ has no zeros in the unit disc. For this reason, analogously to ARMA theory one may prove that the operator $\alpha_{\boldsymbol{\theta}}(B)$ can be inverted. Since $\mathbb{E}|X_0|^{2\eta} < \infty$ for some small enough $\eta > 0$, as will be demonstrated in Example 5.2.5, and since the Taylor coefficients of $\alpha_{\boldsymbol{\theta}}(z)/\beta_{\boldsymbol{\theta}}(z)$ (in the expansion about 0) decay exponentially fast, the difference equation (4.23) has a unique stationary ergodic solution with almost sure representation

$$\begin{aligned} h_t(\boldsymbol{\theta}) &= (\beta_{\boldsymbol{\theta}}(B))^{-1}\alpha_{\boldsymbol{\theta}}(B)\left(\frac{\alpha_0}{\alpha_{\boldsymbol{\theta}}(1)} + X_t^2\right) \\ &= \frac{\alpha_0}{\beta_{\boldsymbol{\theta}}(1)} + (\beta_{\boldsymbol{\theta}}(B))^{-1}\alpha_{\boldsymbol{\theta}}(B)X_t^2 \\ &= \frac{\alpha_0}{\beta_{\boldsymbol{\theta}}(1)} + \sum_{j=1}^{\infty}\pi_j(\boldsymbol{\theta})X_{t-j}^2, \qquad \boldsymbol{\theta} \in K, \end{aligned} \tag{4.26}$$

where the sequence $(\pi_j(\boldsymbol{\theta}))_{j\geq 1}$ is determined by

$$\sum_{j=1}^{\infty} \pi_j(\boldsymbol{\theta}) z^j = \frac{\alpha_{\boldsymbol{\theta}}(z)}{\beta_{\boldsymbol{\theta}}(z)}, \qquad |z| \leq 1.$$

Also observe that $h_t(\boldsymbol{\theta}_0) = \sigma_t^2$ a.s., in which case (4.26) can be interpreted as the ARCH(∞) representation of GARCH(p,q); cf. Giraitis et al [55]. By means of the Cauchy inequalities one can show that there exist $c > 0$ and $0 < \rho < 1$ such that the ith derivatives of π_j with respect to $\boldsymbol{\theta}$ fulfill

$$|\pi_j^{(i)}(\boldsymbol{\theta})| \leq c\rho^j, \qquad j \geq 0, \tag{4.27}$$

for every $\boldsymbol{\theta} \in K$, $i = 0,1,2$. This is utilized to demonstrate that the random functions $h_t(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in K$, $t \in \mathbb{Z}$, are almost surely twice continuously differentiable (in the interior of K); this implies that the matrices $\mathbf{G}_0$ and $\mathbf{F}_0$ of Theorem 4.2.1 are well–defined. By induction it is straightforward to verify that

$$\hat{h}_t(\boldsymbol{\theta}) = \frac{\alpha_0}{\beta_{\boldsymbol{\theta}}(1)} + \sum_{j=1}^{t-1} \pi_j(\boldsymbol{\theta}) X_{t-j}^2 \tag{4.28}$$

for all $t \geq 1$. Thus by combining (4.26)–(4.28), for all $\boldsymbol{\theta} \in K$

$$\begin{aligned} |\hat{h}_t(\boldsymbol{\theta}) - h_t(\boldsymbol{\theta})| &\leq \sum_{j=t}^{\infty} |\pi_j(\boldsymbol{\theta})| X_{t-j}^2 \leq c \sum_{j=t}^{\infty} \rho^j X_{t-j}^2 \\ &= c\rho^t \sum_{k=0}^{\infty} \rho^k X_{-k}^2. \end{aligned} \tag{4.29}$$

Since $\mathbb{E}|X_0|^{2\eta} < \infty$ for $\eta > 0$ small enough, as will be shown in Example 5.2.5), also $\mathbb{E}(\log^+ X_0^2) < \infty$. Therefore from an application of Proposition 2.5.1 to the right–hand side of (4.29),

$$\|\hat{h}_t - h_t\|_K \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty, \tag{4.30}$$

where $\|\cdot\|_K$ denotes the sup–norm in $\mathbb{C}(K)$, the space of continuous functions on K. For this reason the sequence $(h_t)_{t\in\mathbb{N}}$ can be regarded as a "stationary approximation" of $(\hat{h}_t)_{t\in\mathbb{N}}$. This compares with Section 5.2.3. By replacing $\hat{h}_t$, one obtains the random elements L_n, see (4.25). A Taylor argument shows that $n^{-1}\|\hat{L}_n - L_n\|_K \xrightarrow{\text{a.s.}} 0$.

The sequence $(L_n)_{n\in\mathbb{N}}$ is easier to handle than $(\hat{L}_n)_{n\in\mathbb{N}}$ because $(X_t^2/h_t + \log h_t)_{t\in\mathbb{N}}$ is a stationary ergodic sequence. It can be shown that the maximizer of L_n is asymptotically equivalent to $\hat{\boldsymbol{\theta}}_n$; to this end one verifies that $n^{-1}\|\hat{L}_n - L_n\|_K \xrightarrow{\text{a.s.}} 0$ and $n^{-1/2}\|\hat{L}_n' - L_n'\|_K \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. The analysis of the maximizer of L_n then follows from the standard steps of proof for consistency and asymptotic normality of M–estimators.

Observe the "sandwich"–form of the asymptotic covariance matrix $\mathbf{F}_0^{-1}\mathbf{G}_0\mathbf{F}_0^{-1}$. If Z_0 is actually standard Gaussian, then $\mathbf{G}_0 = -\mathbf{F}_0$. Such a structure for the asymptotic covariance matrix is typical for asymptotically normal QMLEs; see e.g. Gouriéroux and Monfort [59] for further examples. This form will also become apparent in our derivation of asymptotic normality of the QMLE in general conditionally heteroscedastic time series models; see Theorem 5.6.1 of Chapter 5.

Some Remarks Concerning the Practical Application of Theorem 4.2.1

To the best of our knowledge it is impossible to explicitly compute the matrices $\mathbf{G}_0$ and $\mathbf{F}_0$. If one wants to determine $\mathbf{G}_0$ or $\mathbf{F}_0$, one has to rely on simulation techniques. In contrast to the asymptotic covariance matrix $[\mathbf{W}(\boldsymbol{\theta}_0)]^{-1}$ of the QMLE in ARMA, see Theorem 4.1.1, the asymptotic covariance matrix $\mathbf{F}_0^{-1}\mathbf{G}_0\mathbf{F}_0^{-1}$ of GARCH(p,q) critically depends on the form of the distribution of Z_0. This statement can be justified by simulations. There are strongly consistent estimators of the matrices $\mathbf{G}_0$ and $\mathbf{F}_0$. Indeed,

$$\hat{\mathbf{A}}_n = \frac{1}{n}(\hat{L}_n'(\hat{\boldsymbol{\theta}}_n))^T\hat{L}_n'(\hat{\boldsymbol{\theta}}_n) \xrightarrow{\text{a.s.}} \mathbf{G}_0, \qquad n \to \infty,$$

$$\hat{\mathbf{B}}_n = \frac{1}{n}\hat{L}_n''(\hat{\boldsymbol{\theta}}_n) \xrightarrow{\text{a.s.}} \mathbf{F}_0, \qquad n \to \infty.$$

See Remark 5.6.2 for an alternative covariance estimator, which avoids the computation of second derivatives of $\hat{L}_n$.

The practical implementation of the QMLE may pose numerical problems. For GARCH$(1,1)$, Zumbach [137] suggests a parameter transform which facilitates the numerical maximization. Furthermore one has to keep in mind that Theorem 4.2.1 is merely an asymptotic result. The small sample behavior of $\hat{\boldsymbol{\theta}}_n$ also depends on the initialization for $X_0^2, \ldots, X_{-p+1}^2$ and $\sigma_0^2, \ldots, \sigma_{-q+1}^2$. We are not aware of any method which would lead to an “optimal” choice of an initialization.

4.2.2 Whittle Estimation

It has already been observed shortly after the introduction of the GARCH model (Bollerslev [18]) that the squares (X_t^2) of a GARCH(p,q) process *formally* obey an ARMA(k,q) equation with $k = \max(p,q)$. Indeed, one can rephrase the defining GARCH equations

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \end{aligned} \tag{4.31}$$

as follows (substitute σ_{t-j}^2 by the equivalent expression $X_{t-j}^2 - (X_{t-j}^2 - \sigma_{t-j}^2)$):

$$X_t^2 = \alpha_0 + \sum_{i=1}^{k} \varphi_i X_{t-i}^2 + \nu_t - \sum_{j=1}^{q} \beta_j \nu_{t-j}, \qquad t \in \mathbb{Z}, \tag{4.32}$$

where

$$\varphi_i = \alpha_i + \beta_i,$$

(if $i \in (p,q]$, set $\alpha_i = 0$ and if $i \in (q,p]$ set $\beta_i = 0$) and

$$\nu_t = X_t^2 - \sigma_t^2 = \sigma_t^2(Z_t^2 - 1), \qquad t \in \mathbb{Z}.$$

It is readily verified that under the assumptions of (σ_t^2) being strictly stationary and $\mathbb{E}\sigma_0^4 < \infty$, the sequence (ν_t) constitutes a (non–iid) white noise sequence (i.e., (ν_t) has mean zero, constant variance and is uncorrelated).

The ARMA structure (4.32) justifies the application of the Whittle estimator (or the least–squares estimator) to (X_t^2), giving estimates of the parameters φ_i and β_j. As a matter of fact, the econometrics folklore on estimation in GARCH processes claims that the least–squares estimator is "inefficient" and should be avoided for this reason, often by referring to the original article of Engle [46], where the ARCH(p) model was introduced. In fact, Engle deals with estimation in a so–called ARCH regression model, which is basically a *static* regression model with ARCH errors. He computes the efficiency gain for the *regression parameter* in case one accounts for the ARCH structure of the errors. Engle's theory is therefore far from being a comparison between least–squares estimator and QMLE of ARCH parameters.

The claimed "inefficiency" of the least–squares estimator, which can e.g. be observed in simulations, is more related to its unfavorable limit properties in the case of $\mathbb{E}X_0^8 = \infty$. Ould Ahmedou Voffal [109] was the first to observe that the least–squares estimator in ARCH(p) converges at a rate slower than $\sqrt{n}$ in case that $\mathbb{E}X_0^8 = \infty$. Giraitis and Robinson [56] provided the theory of the asymptotically equivalent Whittle estimator in the general GARCH(p,q) model, under the assumption that $\mathbb{E}X_0^8 < \infty$. Mikosch and Straumann [102] completed the picture by studying the limit properties of the Whittle estimator in heavy–tailed GARCH(1, 1) processes (i.e., in the case $\mathbb{E}X_0^8 = \infty$); see Chapter 8.

In order to give an exact definition of the Whittle estimator, introduce the polynomial

$$\varphi(z) = \beta(z) - \alpha(z) = 1 - \varphi_1 z - \cdots - \varphi_k z^k,$$

and note that $\alpha(z) = \alpha_1 z + \cdots + \alpha_p z^p$ and $\beta(z) = 1 - \beta_1 z - \cdots - \beta_q z^q$ have already been defined in Section 4.2.1. In backshift operator notation the difference equation (4.32) then becomes

$$\varphi(B)\left(X_t^2 - \frac{\alpha_0}{\varphi(1)}\right) = \beta(B)\nu_t. \tag{4.33}$$

By recalling the criterion (3.23) and formula (3.24), we recognize that the quotient $\alpha_0/\varphi(1)$ in (4.33) is well–defined and equals $\mathbb{E}X_0^2$. Therefore (X_t^2)

is *formally* an ARMA(k,q) process with mean $\alpha_0/\varphi(1)$ and white noise sequence (ν_t); it is not ARMA in the sense of our definition of Section 3.2 because (ν_t) does *not* constitute an iid sequence. The equivalence between (4.31) and (4.33) is useful for determining the second–order moment structure of a GARCH(p,q) process. In particular, we may compute that $(X_t^2 - \mathbb{E}X_0^2)$ has spectral density

$$f(\lambda;\boldsymbol{\vartheta}) = \frac{\sigma_\nu^2}{2\pi}\frac{|\beta(e^{-i\lambda})|^2}{|\varphi(e^{-i\lambda})|^2} = \frac{\sigma_\nu^2}{2\pi}\,g(\lambda;\boldsymbol{\vartheta})$$

by means of relation (4.13). Here, we denote $\sigma_\nu^2 = \mathbb{E}\nu_0^2$ and let $\boldsymbol{\vartheta} = (\varphi_1,\ldots,\varphi_k,\beta_1,\ldots,\beta_q)^T$. The analogy to Section 4.1.3 motivates to define the statistic

$$\bar{\sigma}_{n,X^2}^2(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_j \frac{I_{n,X^2}(\lambda_j)}{g(\lambda_j;\boldsymbol{\vartheta})},$$

where $I_{n,X^2}(\lambda)$ denotes the periodogram of $X_1^2,\ldots,X_n^2$ as defined in (4.15) and where the summation is as usual taken over all Fourier frequencies (4.16). For fixed constants $\bar{\beta} < 1$ and $\bar{\varphi} < 1$, we define the compact set

$$\tilde{K} = \left\{\boldsymbol{\vartheta} \in \mathbb{R}^{k+q} \mid 0 \le \varphi_i \le \bar{\varphi},\ 0 \le \beta_j \le \varphi_j,\ \beta_1 + \cdots + \beta_q \le \bar{\beta}\right\}. \quad (4.34)$$

Then the Whittle estimator of $\boldsymbol{\vartheta}$ is given by

$$\hat{\boldsymbol{\vartheta}}_n = \underset{\boldsymbol{\vartheta}\in\tilde{K}}{\operatorname{argmin}}\ \bar{\sigma}_{n,X^2}^2(\boldsymbol{\vartheta}). \quad (4.35)$$

The asymptotic properties of $\hat{\boldsymbol{\vartheta}}_n$ have been studied in Giraitis and Robinson [56].

Theorem 4.2.3 (Giraitis and Robinson [56]). *Let (X_t) be a stationary* GARCH(p,q) *process with true parameter $\boldsymbol{\theta}_0 = (\alpha_0^\circ,\ldots,\alpha_p^\circ,\beta_1^\circ,\ldots,\beta_q^\circ)^T$, such that conditions* Q.1 *and* Q.2 *of Section 4.2.1 hold.*

(A)*If $\mathbb{E}X_0^4 < \infty$, then the transformed parameter $\boldsymbol{\vartheta}_0 = (\varphi_1^\circ,\ldots,\varphi_k^\circ,\beta_1^\circ,\ldots,\beta_q^\circ)^T$ is consistently estimated by $\hat{\boldsymbol{\vartheta}}_n$, i.e.,*

$$\hat{\boldsymbol{\vartheta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\vartheta}_0, \qquad n \to \infty,$$

and

$$\bar{\sigma}_{n,X^2}^2(\hat{\boldsymbol{\vartheta}}_n) \xrightarrow{\text{a.s.}} \sigma_\nu^2, \qquad n \to \infty.$$

(B)*If moreover $\mathbb{E}X_0^8 < \infty$ and $\boldsymbol{\vartheta}_0$ lies in the interior of $\tilde{K}$, then*

$$\sqrt{n}\,(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{\text{d}} \mathcal{N}\Big(\mathbf{0},\ [\mathbf{W}(\boldsymbol{\vartheta}_0)]^{-1} + [\mathbf{W}(\boldsymbol{\vartheta}_0)]^{-1}\mathbf{V}(\boldsymbol{\vartheta}_0)[\mathbf{W}(\boldsymbol{\vartheta}_0)]^{-1}\Big),$$

where the matrix

$$\mathbf{W}(\boldsymbol{\vartheta}_0) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{\partial \log g(\lambda;\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right)^T\left(\frac{\partial \log g(\lambda;\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\right)\mathrm{d}\lambda$$

is as in Theorem 4.1.1, and

$$\begin{aligned}&\mathbf{V}(\boldsymbol{\vartheta}_0)\\ &= \frac{\pi}{2\sigma^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left(\frac{\partial(1/g(\lambda;\boldsymbol{\vartheta}_0))}{\partial\boldsymbol{\vartheta}}\right)^T\left(\frac{\partial(1/g(\lambda;\boldsymbol{\vartheta}_0))}{\partial\boldsymbol{\vartheta}}\right) f(\lambda,-\omega,\omega)\,\mathrm{d}\lambda\,\mathrm{d}\omega\,.\end{aligned}$$

The function f under the latter integral is the fourth–order cumulant spectrum

$$\begin{aligned}&f(\lambda,\omega,\mu)\\ &= \frac{1}{(2\pi)^3}\sum_{j,k,l\in\mathbb{Z}}\mathrm{Cov}(X_0X_j,\,X_kX_l)\,e^{-i(j\lambda+k\omega+l\mu)},\quad \lambda,\omega,\mu\in(-\pi,\pi].\end{aligned}$$

Remarks 4.2.4.

1. The nonnegative definite matrix $[\mathbf{W}(\boldsymbol{\vartheta}_0)]^{-1}\mathbf{V}(\boldsymbol{\vartheta}_0)[\mathbf{W}(\boldsymbol{\vartheta}_0)]^{-1}$ appearing in the asymptotic covariance matrix of $\hat{\boldsymbol{\vartheta}}_n$ can be seen as an extra term due to the non–linearity of $(X_t^2 - \mathbb{E}X_0^2)$.

2. We are not aware of any method which would yield a tractable expression for the matrix $\mathbf{V}(\boldsymbol{\vartheta}_0)$, and hence we cannot explicitly determine the asymptotic covariance matrix. It is however possible to consistently estimate $\mathbf{V}(\boldsymbol{\vartheta}_0)$ from data, see Remark 2.2 in Giraitis and Robinson [56].

3. Giraitis and Robinson [56] leave out the estimation of the parameter α_0. Estimation of α_0 could be based on the formula

   $$\mathrm{Var}(X_0) = \frac{\alpha_0}{\varphi(1)}\,.$$

   A natural estimator of α_0 is therefore given by

   $$\hat{\alpha}_0 = \gamma_{n,X}(0)(1-\hat{\varphi}_1-\cdots-\hat{\varphi}_k)$$

   where $\hat{\varphi}_j$ is the Whittle estimator of φ_j, $j=1,\ldots,k$. It is possible to determine the joint limit behavior of $(\hat{\alpha}_0,\hat{\alpha}_1,\ldots,\hat{\alpha}_p,\hat{\beta}_1,\ldots,\hat{\beta}_q)^T$. In the light of the fact that the Whittle estimator should not be used for the estimation of GARCH parameters, a detailed elaboration is however not worth doing. Compare with the similar Remark 8.3.7.

4. The moment conditions $\mathbb{E}X_0^4<\infty$ and $\mathbb{E}X_0^8<\infty$ are rather severe, particularly in the light of the fact that financial log–return data often seem to have an infinite 8th moment; see e.g. Chapter 6 of Embrechts et al. [45]. Theorem 4.2.3 does not give any information about the rate of

convergence of $\hat{\boldsymbol{\vartheta}}_n$ in the case of $\mathbb{E}X_0^8 = \infty$ and about consistency when $\mathbb{E}X_0^4 = \infty$. In Chapter 8 we provide a detailed study of the asymptotic properties of $\hat{\boldsymbol{\vartheta}}_n$ in GARCH(1,1) with $\mathbb{E}X_0^8 < \infty$. For the case of a general GARCH(p,q) process, we make the following conjecture concerning the asymptotics of $\hat{\boldsymbol{\vartheta}}_n$: Suppose the distribution of $|X_0|$ has tail index $\kappa > 0$, i.e.,

$$\mathbb{P}(|X_0| > x) \sim cx^{-\kappa}, \qquad x \to \infty, \tag{4.36}$$

where $c > 0$ is some constant. The value of κ governs the asymptotics of $\hat{\boldsymbol{\vartheta}}_n$. As regards the tail index κ, observe that (4.36) implies that $|X_0|$ has a finite pth moment if and only if $p < \kappa$. One has to distinguish between the following cases:

- $\kappa < 4$: The Whittle estimator $\hat{\boldsymbol{\vartheta}}_n$ is inconsistent.

- $4 < \kappa < 8$: The Whittle estimator $\hat{\boldsymbol{\vartheta}}_n$ is strongly consistent, but the rate of convergence is $n^{1-4/\kappa}$, i.e., slower than $\sqrt{n}$. The limit distribution is non–Gaussian.

- $\kappa > 8$: Since $\mathbb{E}X_0^8 < \infty$ is implied, this case is covered by Theorem 4.2.3; one has strong consistency and asymptotic normality.

□

5

Quasi Maximum Likelihood Estimation in Conditionally Heteroscedastic Time Series Models: A Stochastic Recurrence Equations Approach

In this chapter we study the (Gaussian) quasi maximum likelihood estimator (QMLE) in a general conditionally heteroscedastic time series model of multiplicative form

$$\begin{cases} X_t &= \sigma_t Z_t, \\ \sigma_t^2 &= g_{\boldsymbol{\theta}}(X_{t-1}, \ldots, X_{t-p}, \sigma_{t-1}^2, \ldots, \sigma_{t-q}^2), \end{cases} \qquad t \in \mathbb{Z}, \tag{5.1}$$

where the volatility process (σ_t) is nonnegative and (Z_t) is a sequence of iid random variables with $\mathbb{E}Z_0 = 0$ and $\mathbb{E}Z_0^2 = 1$. The parametric family $\{g_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \Theta\}$ of nonnegative functions on $\mathbb{R} \times [0, \infty)$ fulfills certain regularity conditions. We suppose that $\Theta \subset \mathbb{R}^d$. Since we also require that σ_t is $\mathcal{F}_{t-1}$–measurable, the model (5.1) belongs to the previously introduced class (3.6).

The techniques of stochastic recurrence equations (SREs) introduced in Section 2.6 enable us to develop a unifying theory for the QMLE in model (5.1). We will apply this theory to the conditionally heteroscedastic time series models of Section 3.3: AGARCH(p, q) and EGARCH. The contents of this chapter are based on Straumann and Mikosch [125].

5.1 Overview

In this section we merely sketch the main problems and ideas in order to give a flavor of this chapter of the monograph. First of all, we must investigate whether the equations (5.1) admit a stationary solution $((X_t, \sigma_t))$. Substituting $\sigma_{t-i} Z_{t-i}$ for X_{t-i} in the second equation of (5.1) results in

$$\sigma_t^2 = g_{\boldsymbol{\theta}}(\sigma_{t-1} Z_{t-1}, \ldots, \sigma_{t-p} Z_{t-p}, \sigma_{t-1}^2, \ldots, \sigma_{t-q}^2), \qquad t \in \mathbb{Z}, \tag{5.2}$$

and shows that the vectors ${}_k\boldsymbol{\sigma}_t^2 = (\sigma_t^2, \ldots, \sigma_{t-k+1}^2)^T$ with $k = \max(p, q)$ obey a SRE ${}_k\boldsymbol{\sigma}_{t+1}^2 = \psi_t({}_k\boldsymbol{\sigma}_t^2)$ on $[0, \infty)^k$; see (5.6) for the exact definition of (ψ_t).

Stationarity is equivalent to the existence of a nonnegative stationary solution to the latter SRE.

The second important issue is the reconstruction of the unobservable volatility sequence (σ_t) from data. In the context of model (5.1), we introduce the notation

$$\mathbf{X}_t = (X_t, \ldots, X_{t-p+1})^T \quad \text{and} \quad \boldsymbol{\sigma}_t^2 = (\sigma_t^2, \ldots, \sigma_{t-q+1}^2)^T,$$

and write $\sigma_t^2 = g_{\boldsymbol{\theta}}(\mathbf{X}_{t-1}, \boldsymbol{\sigma}_{t-1}^2)$ for short. A natural way to approximate the unobservable vector $\boldsymbol{\sigma}_t^2$ of squared volatilities from data $X_{-p+1}, \ldots, X_n$ is as follows:

1. Set $\hat{\boldsymbol{\sigma}}_0^2 = \varsigma_0^2$ for an arbitrary initial value $\varsigma_0^2 \in [0, \infty)^q$.

2. Define $\hat{\boldsymbol{\sigma}}_{t+1}^2 = \phi_t(\mathbf{X}_t, \hat{\boldsymbol{\sigma}}_t^2)$ for $1 \le t \le n$, where

$$\phi_t(\mathbf{s}) = (g_{\boldsymbol{\theta}_0}(\mathbf{X}_t, \mathbf{s}), s_1, \ldots, s_{q-1})^T, \qquad \mathbf{s} = (s_1, \ldots, s_q)^T \in [0, \infty)^q.$$

Here $\boldsymbol{\theta}_0$ denotes the true parameter. The assumption that $X_0, \ldots, X_{-p+1}$ *are* observed is in contrast to Section 4.2.1, but it may be made for the ease of argument since it turns out to be irrelevant for the asymptotic behavior of the QMLE. One would like that

$$|\hat{\boldsymbol{\sigma}}_t^2 - \boldsymbol{\sigma}_t^2| \xrightarrow{\mathbb{P}} 0, \qquad t \to \infty, \tag{5.3}$$

irrespective of the chosen initialization ς_0^2. In Section 5.2.2 we will explain in more details why property (5.3) has to be regarded as an invertibility condition for the nonlinear model (5.1).

Analogously to the GARCH(p, q) case discussed in Section 4.2.1, we have to define an estimate $\hat{\mathbf{h}}_t(\boldsymbol{\theta})$ of the squared volatility under parameter hypothesis $\boldsymbol{\theta}$. We write $\hat{\mathbf{h}}_t(\boldsymbol{\theta}) = (\hat{h}_t(\boldsymbol{\theta}), \ldots, \hat{h}_{t-q+1}(\boldsymbol{\theta}))^T$ and recognize that it is natural to proceed according to the following recursive scheme for the determination of $(\hat{\mathbf{h}}_t(\boldsymbol{\theta}))_{t\in\mathbb{N}}$:

$$\hat{\mathbf{h}}_t(\boldsymbol{\theta}) = \begin{cases} \varsigma_0^2, & t = 0, \\ \Phi_{t-1}(\hat{\mathbf{h}}_{t-1}(\boldsymbol{\theta})), & t \ge 1, \end{cases}$$

where $[\Phi_t(\mathbf{s})](\boldsymbol{\theta}) = (g_{\boldsymbol{\theta}}(\mathbf{X}_t, \mathbf{s}(\boldsymbol{\theta})), s_1(\boldsymbol{\theta}), \ldots, s_{q-1}(\boldsymbol{\theta}))^T$ for any vector function $\mathbf{s} = (s_1, \ldots, s_q)^T$. Then by the same considerations as in Section 4.2.1, one can define an *approximate* conditional Gaussian log–likelihood by

$$\hat{L}_n(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^{n} \left(\frac{X_t^2}{\hat{h}_t(\boldsymbol{\theta})} + \log \hat{h}_t(\boldsymbol{\theta}) \right).$$

The QMLE $\hat{\boldsymbol{\theta}}_n$ is a maximizer of $\hat{L}_n(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta} \in K$, where K is a suitable compact subset of the parameter space Θ. For analyzing the QMLE, it is essential that we understand the limit behavior of $(\hat{L}_n)$, and hence also of the sequence $(\hat{h}_t)_{t\in\mathbb{N}}$. We desire to approximate $(\hat{\mathbf{h}}_t)_{t\in\mathbb{N}}$ (and thus also $(\hat{h}_t)_{t\in\mathbb{N}}$) by a stationary ergodic sequence $(\mathbf{h}_t)_{t\in\mathbb{N}} = ((h_t,\ldots,h_{t-q+1})^T)_{t\in\mathbb{N}}$ because this simplifies the analysis. It is intuitively clear that $(\mathbf{h}_t)$ obeys the SRE

$$\mathbf{h}_{t+1} = \Phi_t(\mathbf{h}_t), \qquad t \in \mathbb{Z}, \tag{5.4}$$

Hence the problem consists of giving conditions which guarantee the existence of a stationary solution to (5.4). The assumption that Φ_t is a "contraction on average" will be sufficient. We then set

$$L_n(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{n}\left(\frac{X_t^2}{h_t(\boldsymbol{\theta})} + \log h_t(\boldsymbol{\theta})\right),$$

and observe that $n^{-1}\hat{L}_n$ and $n^{-1}L_n$ have the same almost sure limit. For this reason, consistency of $\hat{\boldsymbol{\theta}}_n$ is equivalent to the consistency of the maximizer $\tilde{\boldsymbol{\theta}}_n$ of L_n. Most technicalities for establishing consistency of $\hat{\boldsymbol{\theta}}_n$ emerge from proving the existence of the sequence $(\mathbf{h}_t)$, the remaining arguments are almost routine. Already the verification of invertibility, which is necessary for the existence of an approximating sequence $(\mathbf{h}_t)_{t\in\mathbb{N}}$ since $(\hat{\mathbf{h}}_t(\boldsymbol{\theta}_0))_{t\in\mathbb{N}} = (\hat{\boldsymbol{\sigma}}_t^2)_{t\in\mathbb{N}}$, may be an intricate and unpleasant problem. We are e.g. not able to check invertibility in the general EGARCH model without a simulation approach, see Example 5.2.8.

The step from consistency to asymptotic normality necessitates that we analyze the first and second derivative of L_n with respect to the parameter $\boldsymbol{\theta}$, and hence that we establish differentiability of h_t. For this problem, the SRE approach is again helpful and leads to an elegant unifying theory.

The novel approach presented in this chapter is to *explicitly* formulate and solve the arising problems of quasi maximum likelihood estimation by making use of SREs. In this way a high degree of simplicity and generality can be achieved. A further novelty is the consequent application of the ergodic theorem for $\mathbb{C}(K,\mathbb{R}^{d'})$ valued random elements (Theorem 2.2.1). This tool allows us to avoid the computation of *third* derivatives of the log–likelihood, which leads to more elegant proofs and regularity conditions which are easier to verify. In statistical literature, proofs of the asymptotic normality of M–estimators often impose the classical regularity conditions by Cramér [32], which involve third derivatives.

5.2 Stationarity, Ergodicity and Invertibility

We recall that

$$\mathcal{F}_t = \sigma(Z_j,\, j \le t)$$

denotes the σ–field generated by the random variables $\{Z_j \mid j \le t\}$ and appeal to Section 2.6 for the techniques of SREs.

5.2.1 Existence of a Stationary Solution

In this and the following section we suppress the parameter $\boldsymbol{\theta}$ in our notation because for the treatment of stationarity and invertibility it can be assumed as fixed. In order to discuss the stationarity issue, we first embed model (5.1) into a SRE. To this end, introduce ${}_k\boldsymbol{\sigma}_t^2 = (\sigma_t^2, \ldots, \sigma_{t-k+1}^2)^T$ and set ${}_k\mathbf{Z}_t = (Z_t, \ldots, Z_{t-k+1})^T$ for $k = \max(p, q)$. Then by substituting the X_{t-i}'s by $Z_{t-i}\sigma_{t-i}$ in the second equation of (5.1), we observe that $({}_k\boldsymbol{\sigma}_t^2)$ is a solution of the SRE

$$\mathbf{s}_{t+1} = \psi_t(\mathbf{s}_t), \qquad t \in \mathbb{Z}, \tag{5.5}$$

on $[0, \infty)^k$, where

$$\psi_t(\mathbf{s}) = (g(\mathbf{s}^{1/2} \odot {}_k\mathbf{Z}_t\, , \mathbf{s}), s_1, \ldots, s_{k-1})^T, \tag{5.6}$$

with $\mathbf{s}^{1/2} = (s_1^{1/2}, \ldots, s_k^{1/2})^T$ for $\mathbf{s} \in [0, \infty)^k$; recall that $\odot$ stands for the Hadamard product, the componentwise multiplication of matrices or vectors of the same dimension. On the other hand, if a stationary sequence $(\mathbf{s}_t)$ is a solution of (5.5) and $\mathbf{s}_t$ is measurable with respect to $\mathcal{F}_{t-1}$, then the sequence $((\mathbf{s}_{t,1}^{1/2} Z_t, \mathbf{s}_{t,1}^{1/2}))$ (here $\mathbf{s}_{t,1}^{1/2}$ denotes the first coordinate of $\mathbf{s}_t^{1/2}$) is stationary and obeys the equations (5.1). Hence for proving the existence and uniqueness of a stationary sequence $((X_t, \sigma_t))$ fulfilling (5.1) and σ_t^2 being $\mathcal{F}_{t-1}$–measurable, we may focus on showing that there exists a unique stationary nonnegative solution $(\mathbf{s}_t)$ to (5.5) for which $\mathbf{s}_t$ is $\mathcal{F}_{t-1}$–measurable. After noticing that (ψ_t) is stationary ergodic, an application of Theorem 2.6.1 with $(\phi_t) = (\psi_t)$ and d the Euclidean metric results in the following proposition.

Proposition 5.2.1. *Fix an arbitrary $\varsigma_0^2 \in [0, \infty)$ and suppose that the following conditions hold true for the stationary ergodic sequence (ψ_t) of maps defined in* (5.6)*:*

S.1 $\mathbb{E}(\log^+ |\psi_0(\varsigma_0^2)|) < \infty$.

S.2 $\mathbb{E}[\log^+ \Lambda(\psi_0)] < \infty$ *and for some integer $r \geq 1$ it holds that* $\mathbb{E}[\log \Lambda(\psi_0^{(r)})] < 0$.

Then the SRE (5.5) *admits a unique stationary ergodic solution $({}_k\boldsymbol{\sigma}_t^2)$ such that ${}_k\boldsymbol{\sigma}_t^2$ is $\mathcal{F}_{t-1}$–measurable for every $t \in \mathbb{Z}$. The following almost sure representation is valid:*

$${}_k\boldsymbol{\sigma}_t^2 = \lim_{m\to\infty} \psi_{t-1} \circ \cdots \circ \psi_{t-m}(\varsigma_0^2), \qquad t \in \mathbb{Z}, \tag{5.7}$$

where the latter limit is irrespective of ς_0^2. For any other solution $({}_k\tilde{\boldsymbol{\sigma}}_t^2)$ of the SRE (5.5) *with index sets $\mathbb{Z}$ or $\mathbb{N}$,*

$$|\,{}_k\tilde{\boldsymbol{\sigma}}_t^2 - {}_k\boldsymbol{\sigma}_t^2\,| \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty. \tag{5.8}$$

Remark 5.2.2. In general, the stationary distribution induced by the SRE (5.5) is not known and thus the perfect simulation of $((X_t, \sigma_t))_{t\in\mathbb{N}}$ is impossible. The following algorithm provides an approximation with an error decaying exponentially fast almost surely:

1. Take an initial value $\varsigma_0^2 \in [0,\infty)^k$, set ${}_k\tilde{\boldsymbol{\sigma}}_0^2 = \varsigma_0^2$, and generate ${}_k\tilde{\boldsymbol{\sigma}}_t^2$ according to (5.5).
2. Set $(\tilde{X}_t, \tilde{\sigma}_t) = \left({}_k\tilde{\boldsymbol{\sigma}}_{t,1}^2\right)^{1/2} (Z_t, 1), \qquad t = 0, 1, \ldots$.

From Proposition 5.2.1 it is immediate that $|\tilde{\sigma}_t^2 - \sigma_t^2| \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$. Moreover, by virtue of Proposition 2.5.1 and Lemma 2.5.4 we can conclude $|\tilde{X}_t - X_t| = |Z_t|\,|\tilde{\sigma}_t - \sigma_t| \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$. Compare with Remarks 2.6.2. □

We now apply Proposition 5.2.1 to AGARCH(1,1) and EGARCH.

Example 5.2.3. The AGARCH(1,1) process of Section 3.3.1 has a squared volatility of the form

$$\sigma_t^2 = \alpha_0 + \alpha_1(|X_{t-1}| - \gamma X_{t-1})^2 + \beta_1\sigma_{t-1}^2, \qquad t \in \mathbb{Z},$$

where $\alpha_0 > 0$, $\alpha_1, \beta_1 \geq 0$ and $|\gamma| \leq 1$. Thus

$$g(x,s) = \alpha_0 + \alpha_1(|x| - \gamma x)^2 + \beta_1 s$$

and

$$\psi_t(s) = g(\sqrt{s}Z_t, s) = \alpha_0 + (\alpha_1(|Z_t| - \gamma Z_t)^2 + \beta_1)s.$$

Note that $\log \Lambda(\psi_0^{(r)}) = \sum_{i=1}^{r} \log(\alpha_1(|Z_{1-i}| - \gamma Z_{1-i})^2 + \beta_1)$. For this reason condition S.2 of Proposition 5.2.1 is equivalent to

$$\mathbb{E}[\log(\alpha_1(|Z_0| - \gamma Z_0)^2 + \beta_1)] < 0. \tag{5.9}$$

Condition (5.9) is also necessary for stationarity as was indicated on p. 54. Since relation (5.7) is valid for arbitrary initial values ς_0^2, we have that

$$\begin{aligned}
\sigma_t^2 &= \lim_{m\to\infty} \psi_{t-1} \circ \cdots \circ \psi_{t-m}(0) \\
&= \lim_{m\to\infty} \alpha_0 \left(1 + \sum_{k=1}^{m-1} \prod_{i=1}^{k} (\alpha_1(|Z_{t-i}| - \gamma Z_{t-i})^2 + \beta_1)\right) \\
&= \alpha_0 \left(1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\alpha_1(|Z_{t-i}| - \gamma Z_{t-i})^2 + \beta_1)\right) \quad \text{a.s.}
\end{aligned}$$

This is once again formula (3.19).

Example 5.2.4. The EGARCH model introduced in Section 3.3.2 has a volatility obeying the SRE

$$\log \sigma_t^2 = \alpha + \beta \log \sigma_{t-1}^2 + \gamma Z_{t-1} + \delta |Z_{t-1}|, \qquad t \in \mathbb{Z}, \tag{5.10}$$

where $\alpha \in \mathbb{R}$, $\delta \geq |\gamma|$ and $0 \leq \beta < 1$. In other words, the sequence $(\log \sigma_t^2)$ constitutes a causal AR(1) process with non–centered innovations sequence $(\alpha + \gamma Z_{t-1} + \delta |Z_{t-1}|)$ or, equivalently, a causal AR(1) process with nonzero mean. Although EGARCH falls into the class of models defined by (5.1), we avoid writing the SRE (5.10) in the form $\sigma_t^2 = g(X_{t-1}, \sigma_{t-1}^2)$ since this would unnecessarily complicate the question of stationarity. Instead we apply Theorem 2.6.1 to the SRE

$$\log \sigma_{t+1}^2 = \psi_t(\log \sigma_t^2), \qquad t \in \mathbb{Z}, \tag{5.11}$$

where

$$\psi_t(s) = \alpha + \beta s + \gamma Z_t + \delta |Z_t|, \qquad t \in \mathbb{Z}.$$

We recognize that $|\beta| < 1$ together with $\mathbb{E}[\log^+(\alpha + \gamma Z_0 + \delta |Z_0|)] < \infty$ is a sufficient condition for the existence of a unique stationary solution to (5.10); since $\mathbb{E} Z_0^2 = 1$, the innovations $(\alpha + \gamma Z_{t-1} + \delta |Z_{t-1}|)$ automatically have a finite positive logarithmic moment by Lemma 2.5.3. By taking the limit of the backward iterates associated with the SRE (5.11), it is straightforward to see that the unique stationary solution $(\log \sigma_t^2)$ of (5.10) has the almost sure representation

$$\log \sigma_t^2 = \alpha (1-\beta)^{-1} + \sum_{k=0}^{\infty} \beta^k \left(\gamma Z_{t-1-k} + \delta |Z_{t-1-k}|\right), \qquad t \in \mathbb{Z}, \tag{5.12}$$

which is well–known from the theory of the causal AR(1) process, see e.g. Brockwell and Davis [29]. □

Example 5.2.5. In the AGARCH(p, q) model of Section 3.3.1 the squared volatility is of form

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i (|X_{t-i}| - \gamma X_{t-i})^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \qquad t \in \mathbb{Z}, \tag{5.13}$$

where $\alpha_0 > 0$, $\alpha_i, \beta_j \geq 0$ and $|\gamma| \leq 1$. Note that $\gamma = 0$ in AGARCH(p, q) corresponds to a GARCH(p, q) process. It is possible to treat stationarity of AGARCH by means of Proposition 5.2.1, but the resulting sufficient conditions would not be very enlightening. Instead one adapts the particular methods developed in Bougerol and Picard [21] for the treatment of GARCH(p, q) and derives a SRE with iid *linear* random transition maps. This was worked out in Section 3.3.1. There we introduced

$$\mathbf{Y}_t = (\sigma_t^2, \ldots, \sigma_{t-q+1}^2, (|X_{t-1}| - \gamma X_{t-1})^2, \ldots, (|X_{t-p+1}| - \gamma X_{t-p+1})^2\,)^T,$$

the $(p+q-1) \times (p+q-1)$–matrix valued iid sequence $(\mathbf{A}_t)$ and the vector $\mathbf{B}$, given by

$$\mathbf{A}_t = \begin{pmatrix} \alpha_1(|Z_t|-\gamma Z_t)^2+\beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q & \alpha_2 & \alpha_3 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ (|Z_t|-\gamma Z_t)^2 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \qquad (5.14)$$

$$\mathbf{B} = (\alpha_0, 0, \cdots, 0)^T \in \mathbb{R}^{p+q-1}.$$

Then $(\mathbf{Y}_t)$ obeys the SRE

$$\mathbf{Y}_{t+1} = \mathbf{A}_t \mathbf{Y}_t + \mathbf{B}, \qquad t \in \mathbb{Z}. \qquad (5.15)$$

We showed that there is a unique stationary AGARCH(p,q) process if and only if the SRE (5.15) admits a unique nonnegative stationary solution. The latter is in turn equivalent to $(\mathbf{A}_t)$ having a negative top Lyapunov exponent, i.e.,

$$\rho = \inf_{t \in \mathbb{N}} \Big\{ \frac{1}{t+1} \mathbb{E}\big(\log \|\mathbf{A}_0 \cdots \mathbf{A}_{-t}\|_{\mathrm{op}} \big) \Big\} < 0. \qquad (5.16)$$

Here $\|\cdot\|_{\mathrm{op}}$ denotes the matrix operator norm (2.9) with respect to the Euclidean norm on $\mathbb{R}^{p+q-1}$. Moreover, $(\mathbf{Y}_t)$ is ergodic and measurable with respect to $\mathcal{F}_{t-1}$ and almost sure representation

$$\mathbf{Y}_t = \sum_{k=0}^{\infty} \Big(\prod_{\ell=1}^{k} \mathbf{A}_{t-\ell} \Big) \mathbf{B}, \qquad t \in \mathbb{Z}. \qquad (5.17)$$

We had that

$$\sum_{i=1}^{p} \alpha_i \, \mathbb{E}[(|Z_0| - \gamma Z_0)^2] + \sum_{j=1}^{q} \beta_j < 1$$

is sufficient for stationarity and implies $\mathbb{E}X_0^2 < \infty$, and we recognized that

$$\sum_{j=1}^{q} \beta_j < 1$$

is necessary for stationarity of AGARCH(p, q). In what follows, we show that

$$\mathbb{E}|\mathbf{Y}_0|^\eta < \infty$$

if η is small enough. By (5.16) there is $r \geq 1$ with $\mathbb{E}(\log \|\mathbf{A}_0 \cdots \mathbf{A}_{-r+1}\|_{\text{op}}) = \mathbb{E}(\log \|\mathbf{A}_0^{(r)}\|_{\text{op}}) < 0$. This and Lemma 2.5.5 imply the existence of $\eta > 0$ and $0 < \lambda < 1$ such that

$$\mathbb{E}\Big[\Big(\prod_{\ell=1}^{m} \|\mathbf{A}^{(r)}_{(1-\ell)r}\|_{\text{op}}\Big)^\eta\Big] = \lambda^m, \qquad m \geq 1. \tag{5.18}$$

Without loss of generality, $\eta \leq 1$. Since every matrix norm is equivalent to the Frobenius norm (2.7), $\mathbb{E}\|\mathbf{A}_0\|_{\text{op}} < \infty$, and hence also $\mathbb{E}\|\mathbf{A}_0\|_{\text{op}}^\eta < \infty$. The identity (5.18) together with the facts that $\|\cdot\|_{\text{op}}$ is submultiplicative and $(\mathbf{A}_t)$ is iid demonstrates

$$\mathbb{E}\|\mathbf{A}_0^{(k)}\|_{\text{op}}^\eta \leq \big(\mathbb{E}\|\mathbf{A}_0^{(r)}\|_{\text{op}}^\eta\big)^{[k/r]}\big(\mathbb{E}\|\mathbf{A}_0\|_{\text{op}}^\eta\big)^{r-r[k/r]} \leq c\lambda^{[k/r]}, \qquad k \geq 0,$$

where $c = \max(1, (\mathbb{E}\|\mathbf{A}_0\|_{\text{op}}^\eta)^r)$. Finally by an application of the Minkowski inequality to (5.17),

$$\mathbb{E}\|\mathbf{Y}_1\|_{\text{op}}^\eta \leq \sum_{k=0}^{\infty} \mathbb{E}\|\mathbf{A}_0^{(k)}\|_{\text{op}}^\eta\, |\mathbf{B}|^\eta \leq c\alpha_0^\eta \sum_{k=0}^{\infty} \lambda^{[k/r]} < \infty.$$

Note also that $\mathbb{E}|\mathbf{Y}_0|^\eta < \infty$ entails

$$\mathbb{E}|\sigma_0|^{2\eta} < \infty.$$

□

5.2.2 Invertibility

For real–life data sets which we assume to be generated by a model of type (5.1), the volatility σ_t will be unobservable. In such a case, it is natural to approximate the unobservable squared volatilities $\boldsymbol{\sigma}_t^2 = (\boldsymbol{\sigma}_t^2, \ldots, \boldsymbol{\sigma}_{t-q+1}^2)^T$ from data $X_{-p+1}, X_{-p+2}, \ldots$ in the following way:

Initialization

1. Set $\hat{\boldsymbol{\sigma}}_0^2 = \varsigma_0^2$, where $\varsigma_0^2 \in [0, \infty)^q$ is an arbitrarily chosen vector.

Recursion

2. Let $\hat{\boldsymbol{\sigma}}_{t+1}^2 = \phi_t(\hat{\boldsymbol{\sigma}}_t^2)$ for $t = 0, 1, 2, \ldots$, where the random functions ϕ_t are defined as follows:

$$\phi_t(\mathbf{s}) = (g(\mathbf{X}_t, \mathbf{s}), s_1, \ldots, s_{q-1})^T, \qquad \mathbf{s} = (s_1, \ldots, s_q)^T \in [0, \infty)^q. \tag{5.19}$$

In the context of our nonlinear model we say that the unique stationary ergodic solution $((X_t, \sigma_t))$ to the equations (5.1) is *invertible* (or: model (5.1) is invertible) if

$$|\hat{\boldsymbol{\sigma}}_t^2 - \boldsymbol{\sigma}_t^2| \xrightarrow{\mathbb{P}} 0, \qquad t \to \infty.$$

Note that this is equivalent to $|\hat{\sigma}_t^2 - \sigma_t^2| \xrightarrow{\mathbb{P}} 0$. In other words, invertibility guarantees that the above algorithm converges. There is a second interpretation, which at the same time clarifies the relationship between invertibility in ARMA models and our notion. Note that $(\hat{\boldsymbol{\sigma}}_t^2)_{t\in\mathbb{N}}$ is a solution of the SRE

$$\mathbf{s}_{t+1} = \phi_t(\mathbf{s}_t), \qquad t \in \mathbb{N}, \tag{5.20}$$

on $[0, \infty)$ and recall that the backward iterates associated with (ϕ_t) are defined by

$$\boldsymbol{\sigma}_{t,0}^2 = \varsigma_0^2,$$

$$\boldsymbol{\sigma}_{t,m}^2 = \phi_{t-1} \circ \cdots \circ \phi_{t-m}(\varsigma_0^2), \qquad m \geq 1,$$

for $t \in \mathbb{Z}$. Observe the relationship $\hat{\boldsymbol{\sigma}}_t^2 = \boldsymbol{\sigma}_{t,t}^2$ for all $t \geq 0$. Furthermore, since we suppose that $((X_t, \sigma_t))$ is stationary, $((\boldsymbol{\sigma}_{t,m}^2, \boldsymbol{\sigma}_t^2))$ is stationary for every fixed $m \geq 0$ (Proposition 2.1.1). Thus

$$\boldsymbol{\sigma}_{t,m}^2 - \boldsymbol{\sigma}_t^2 \stackrel{\mathrm{d}}{=} \boldsymbol{\sigma}_{m,m}^2 - \boldsymbol{\sigma}_m^2 = \hat{\boldsymbol{\sigma}}_m^2 - \boldsymbol{\sigma}_m^2, \qquad m \geq 0.$$

Therefore invertibility is equivalent to

$$\boldsymbol{\sigma}_{t,m}^2 \xrightarrow{\mathbb{P}} \boldsymbol{\sigma}_t^2, \qquad m \to \infty, \tag{5.21}$$

for every $t \in \mathbb{Z}$. Relation (5.21) together with Corollary 2.1.3 and a subsequence argument implies the existence of a measurable function f such that

$$\boldsymbol{\sigma}_t^2 = f(X_{t-1}, X_{t-2}, \ldots) \quad \text{a.s.}$$

for every $t \in \mathbb{Z}$. In other words, given one has all the observations of the past, one can evaluate σ_t^2. If we impose $\sigma_t^2 > 0$ a.s. for the model (5.1), the invertibility allows us to represent Z_t as a function of the past and present observations $\{X_{t-k} \mid k \geq 0\}$. Compare this with the notion of invertibility in ARMA. Recall that an ARMA(p, q) model with parameters $(\varphi_1, \ldots, \varphi_p, \vartheta_1, \ldots, \vartheta_q)^T$ is a stochastic process (X_t) which obeys the difference equation

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{j=1}^{q} \vartheta_j Z_{t-j} + Z_t, \qquad t \in \mathbb{Z},$$

where (Z_t) is a given white noise sequence with mean zero, see Section 3.2. Invertibility is defined as follows: there exists an absolutely summable sequence of constants $(\pi_j)_{j\geq 0}$ such that $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$ a.s., or, in other words, the innovation at time t is a linear functional of the past and present observations $\{X_{t-k} \,|\, k \geq 0\}$. If the two characteristic polynomials $\varphi(z) = 1 - \sum_{i=1}^{p} \varphi_i z^i$ and $\vartheta(z) = 1 + \sum_{j=1}^{q} \vartheta_j z^j$ have no common roots, then the ARMA process is invertible if and only if $\vartheta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. For nonlinear time series models, however, the invertibility issue can be a hard problem. Our notion of invertibility is an adaptation of the notion introduced by Granger and Andersen [61] in the context of a general nonlinear autoregressive moving average model. The following proposition is an immediate consequence of Theorem 2.6.1 applied to the SRE (5.20) with d the Euclidean metric.

Proposition 5.2.6. *Assume that there exists a unique stationary ergodic solution* $((X_t, \sigma_t))$ *to the equations* (5.1). *Suppose in addition for* (ϕ_t) *given in* (5.19)*:*

S.1 $\mathbb{E}(\log^+ |\phi_0(\varsigma_0^2)|) < \infty$.

S.2 $\mathbb{E}[\log^+ \Lambda(\phi_0)] < \infty$ *and for some integer* $r \geq 1$ *it holds that* $\mathbb{E}[\log \Lambda(\phi_0^{(r)})] < 0$.

Then $((X_t, \sigma_t))$ *is invertible. In particular, irrespective of* ς_0^2,

$$|\hat{\boldsymbol{\sigma}}_t^2 - \boldsymbol{\sigma}_t^2| \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty.$$

Moreover we have the stochastic representation

$$\boldsymbol{\sigma}_t^2 = \lim_{m\to\infty} \phi_{t-1}^{(m)}(\varsigma_0^2) \quad \text{a.s.}$$

Example 5.2.7 (Continuation of Example 5.2.3). In the AGARCH$(1,1)$ model, $\phi_t(s) = \alpha_0 + \alpha_1(|X_t| - \gamma X_t)^2 + \beta_1 s$. Condition S.1 is fulfilled by virtue of $\mathbb{E}|X_0|^{2\eta} < \infty$ for $\eta > 0$ small enough (see Example 5.2.5). Since $\Lambda(\phi_0^{(r)}) = \beta_1^r$, the restriction $\beta_1 < 1$ is sufficient for invertibility of AGARCH$(1,1)$. But $\beta_1 < 1$ is already implied by the necessary and sufficient stationarity condition (5.9) since $\log \beta_1 \leq \mathbb{E}[\log(\alpha_1(|Z_0| - \gamma Z_0)^2 + \beta_1)] < 0$. Hence every stationary AGARCH$(1,1)$ process is invertible. By Proposition 5.2.6 we also have the a.s. representation $\sigma_t^2 = \lim_{m\to\infty} \phi_{t-1}^{(m)}(0)$. Thus

$$\sigma_t^2 = \alpha_0(1-\beta_1)^{-1} + \alpha_1 \sum_{k=1}^{\infty} \beta_1^{k-1}(|X_{t-k}| - \gamma X_{t-k})^2 \quad \text{a.s.} \tag{5.22}$$

□

Example 5.2.8 (Continuation of Example 5.2.4). Recall that $\alpha \in \mathbb{R}$, $0 \leq \beta < 1$ and $\delta \geq |\gamma|$ for the parameters of EGARCH. As we mentioned in Example 5.2.4, it is beneficial to consider the SRE for $(\log \sigma_t^2)$, which slightly differs from the general setup of this chapter. From (5.12) and $\gamma z + \delta|z| \geq 0$ for all $z \in \mathbb{R}$ one concludes $\log \sigma_t^2 \geq \alpha(1-\beta)^{-1}$, so that we may interpret (5.10) as a SRE

$$\log \sigma_{t+1}^2 = \phi_t(\log \sigma_t^2), \qquad t \in \mathbb{Z}, \tag{5.23}$$

on the (restricted) set $I = [\alpha(1-\beta)^{-1}, \infty)$, where

$$\phi_t(s) = \alpha + \beta s + (\gamma X_t + \delta |X_t|) \exp(-s/2).$$

Condition S.1 is obviously met because of $\mathbb{E}(\log \sigma_0^2) < \infty$. As regards S.2, let us determine $\Lambda(\phi_0)$. Since ϕ_0 is continuously differentiable, $\Lambda(\phi_0) = \sup_{s \in I} |\phi_0'(s)|$. Maximizing

$$|\phi_0'(s)| = |\beta - 2^{-1}(\gamma X_0 + \delta |X_0|) \exp(-s/2)|$$

over I, we obtain

$$\Lambda(\phi_0) = \max\left(\beta, 2^{-1} \exp(-\alpha(1-\beta)^{-1}/2)(\gamma X_0 + \delta |X_0|) - \beta\right). \tag{5.24}$$

It does not seem to be possible to derive a tractable expression for $\Lambda(\phi_0^{(r)})$ when $r > 1$ and it is not clear how to find a bound sharper than the trivial bound $\Lambda(\phi_0^{(r)}) \le \Lambda(\phi_0) \cdots \Lambda(\phi_{-r+1})$. Recalling the representation (5.12) and substituting X_0 by $\sigma_0 Z_0$ in (5.24), the condition $\mathbb{E}[\log \Lambda(\phi_0)] < 0$, which implies invertibility, reads

$$\mathbb{E}\left[\, \log\left(\max\{\beta, Y_0\}\right) \right] < 0, \tag{5.25}$$

where

$$Y_0 = 2^{-1} \exp\left(2^{-1} \sum_{k=0}^{\infty} \beta^k (\gamma Z_{-k-1} + \delta |Z_{-k-1}|)\right) (\gamma Z_0 + \delta |Z_0|) - \beta.$$

In practice one would have to rely on simulation methods for its verification.

Proposition 5.2.9. *Let* (X_t) *be a stationary ergodic* EGARCH *process. Then the condition* (5.25) *is sufficient for invertibility since*

$$|\hat\sigma_t^2 - \sigma_t^2| \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty. \tag{5.26}$$

Proof. By an application of Theorem 2.6.1 to the SRE (5.23) one has $|\log \hat\sigma_t^2 - \log \sigma_t^2| \xrightarrow{\text{e.a.s.}} 0$. Taking the expectation in (5.12) and accounting for $\mathbb{E}|Z_t| \le (\mathbb{E}Z_t^2)^{1/2} = 1$, we obtain $\mathbb{E}(\log \sigma_0^2) \le (\alpha + \delta)(1-\beta)^{-1} < \infty$. Consequently (5.26) is valid by virtue of Lemma 2.5.4 (ii). □

Remarks 5.2.10.

1. In contrast to AGARCH(1,1) it seems impossible in EGARCH to find an explicit and accessible representation for σ_t^2 in terms of past observations.

2. The case $\beta = 0$ leads to a simpler condition and illustrates that there exist invertible EGARCH models. For $\beta = 0$ the condition (5.25) becomes $-\log 2 + (\delta/2)\mathbb{E}|Z_0| + \mathbb{E}[\log(\gamma Z_0 + \delta|Z_0|)] < 0$. Note that $\delta \le 1$ implies the latter condition. Indeed, since $\mathbb{E}|Z_0| \le (\mathbb{E}Z_0^2)^{1/2} = 1$ we have $\mathbb{E}(\gamma Z_0 + \delta|Z_0|) \le \delta \le 1$. Now the Jensen inequality implies $\mathbb{E}[\log(\gamma Z_0 + \delta|Z_0|)] \le 0$. From the relation $1/2 < \log 2$ one obtains $(\delta/2)\mathbb{E}|Z_0| - \log 2 < 0$, which proves the assertion.

3. We restricted the SRE (5.23) to the set $I = [\alpha(1-\beta)^{-1}, \infty)$. This is not necessary, as can be seem from the arguments in Section 5.4.1. □

Example 5.2.11 (Continuation of Example 5.2.5). For the AGARCH(p,q) model the random maps (5.19) are affine linear, i.e.,

$$\phi_t(\mathbf{s}) = \Big(\alpha_0 + \sum_{i=1}^{p} \alpha_i(|X_{t+1-i}| - \gamma X_{t+1-i})^2\Big)\mathbf{e}_1 + \mathbf{C}\mathbf{s}, \qquad \mathbf{s} \in [0,\infty)^q,$$

where

$$\mathbf{e}_1 = (1,0,\dots,0)^T \in \mathbb{R}^q,$$

$$\mathbf{C} = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \cdots & \beta_{q-1} & \beta_q \\ 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 1 & 0 \end{pmatrix} \in \mathbb{R}^{q\times q}.$$

Condition S.1 is fulfilled by virtue of $\mathbb{E}|X_0|^{2\eta} < \infty$ for $\eta > 0$ small enough (see Example 5.2.5). Regarding S.2, observe that $\Lambda(\phi_0^{(r)}) \le \|\mathbf{C}^r\|_{\text{op}}$ for any $r \in \mathbb{N}$ (the inequality in the latter relation is a consequence of the fact that the domain of ϕ_t is but a subset of $\mathbb{R}^q$). In order to prove $\log\|\mathbf{C}^r\|_{\text{op}} < 0$ for large enough r, first recall that a necessary condition for stationarity in AGARCH(p,q) is $\beta_\Sigma := \sum_{j=1}^q \beta_j < 1$. Arguing by recursion on q and expanding the determinant with respect to the last column, it can be shown that $\mathbf{C}$ has characteristic polynomial $p(\lambda) = \det(\lambda\mathbf{I}_q - \mathbf{C}) = \lambda^q(1 - \sum_{j=1}^q \beta_j\lambda^{-i})$. If $|\lambda| > \beta_\Sigma^{1/q}$, then by repeated application of the the triangle inequality together with $0 < \beta_\Sigma < 1$,

$$|p(\lambda)| \ge 1 - \sum_{j=1}^q \beta_j|\lambda|^{-j} > 1 - \sum_{j=1}^q \beta_j\beta_\Sigma^{-j/q} \ge 1 - \beta_\Sigma^{-1}\sum_{j=1}^q \beta_j = 0.$$

Consequently, the matrix $\mathbf{C}$ has spectral radius strictly smaller than $\beta_\Sigma^{1/q}$. By the Jordan normal decomposition this entails

$$\|\mathbf{C}^r\|_{\text{op}} \le c\beta_\Sigma^{r/q}, \qquad r \ge 0,$$

where c is a constant depending on q and β_Σ, and thus $\log \Lambda(\phi_0^{(r)}) \to -\infty$ as $r \to \infty$. This means that the conditions of Proposition 5.2.6 are met for stationary AGARCH(p,q) processes. Consequently, *every* stationary AGARCH(p,q) process is automatically invertible. It is also possible to give an explicit representation of σ_t^2 in terms of $(X_{t-1}, X_{t-2}, \dots,)$; see equation (5.43) below.

Let us also indicate how a bound on $\|\mathbf{C}^r\|_{\text{op}}$ is directly derived. Take an arbitrary $\mathbf{b}_0 \in \mathbb{R}^q$ and let $\mathbf{b}_r = \mathbf{C}^r\mathbf{b}_0$, where $r \in \mathbb{N}$. We claim that

$$|\mathbf{C}^r\mathbf{b}_0| \le \sqrt{q}\,\beta_\Sigma^{[r/q]}\,|\mathbf{b}_0|, \qquad r \in \mathbb{N}. \tag{5.27}$$

Indeed, due to the special structure of $\mathbf{C}$, we have

$$\mathbf{b}_{r,1} = \sum_{j=1}^{q} \beta_j \mathbf{b}_{r-1,j} \quad \text{and} \quad \mathbf{b}_{r,k} = \mathbf{b}_{r-1,k-1}, \quad r \ge 1,\ 2 \le k \le q. \tag{5.28}$$

Since $\max(\,|\mathbf{b}_{r,j}| \mid j = 1, \dots, q) \le |\mathbf{b}_r|$ for all r, relations (5.28) imply $|\mathbf{b}_{r,1}| \le \beta_\Sigma |\mathbf{b}_{r-1}|$. Hence $|\mathbf{b}_{1,1}| \le \beta_\Sigma |\mathbf{b}_0|$. By an inductive argument exploiting (5.28),

$$|\mathbf{b}_{i,1}| \le \beta_\Sigma |\mathbf{b}_0|, \qquad i = 1, \dots, q.$$

Due to (5.28) one has $\mathbf{b}_q = (\mathbf{b}_{q,1}, \mathbf{b}_{q-1,1}, \dots, \mathbf{b}_{1,1})^T$, which implies $|\mathbf{b}_{q,j}| \le \beta_\Sigma |\mathbf{b}_0|$ for all j. Thus $|\mathbf{b}_{q+1,1}| \le \beta_\Sigma^2 |\mathbf{b}_0|$, and so on, yielding $|\mathbf{b}_{r,1}| \le \beta_\Sigma^{[r/q]+1}\,|\mathbf{b}_0|$ for all $r \in \mathbb{N}$. This bound on the coordinates of $\mathbf{b}_r$ gives (5.27), and as a trivial implication

$$\|\mathbf{C}^r\|_{\text{op}} \le \sup_{\mathbf{b}_0 \in \mathbb{R}^q \setminus \{\mathbf{0}\}} |\mathbf{C}^r\mathbf{b}_0| / |\mathbf{b}_0| \le \sqrt{q}\,\beta_\Sigma^{[r/q]} \to 0, \qquad r \to \infty.$$

□

5.2.3 Definition of the Function $\mathbf{h}_t$

In this section we do not suppress the parameter $\boldsymbol{\theta}$ anymore, i.e., we write $\sigma_{t+1}^2 = g_{\boldsymbol{\theta}}(\mathbf{X}_t, \boldsymbol{\sigma}_t^2)$. Assume that $\boldsymbol{\theta}$ belongs to some compact parameter space $K \subset \Theta \subset \mathbb{R}^d$ and that $((X_t, \sigma_t))$ is the unique stationary ergodic solution to the equations (5.1) with true parameter $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. For any initial value $\boldsymbol{\varsigma}_0^2 \in [0,\infty)^q$ we define the following random vector functions $\hat{\mathbf{h}}_t$ on K:

$$\hat{\mathbf{h}}_t = \begin{cases} \boldsymbol{\varsigma}_0^2, & t = 0, \\ \Phi_{t-1}(\hat{\mathbf{h}}_{t-1}), & t \ge 1, \end{cases} \tag{5.29}$$

where the random maps $\Phi_t : \mathbb{C}(K,[0,\infty)^q) \to \mathbb{C}(K,[0,\infty)^q)$ are given by

$$[\Phi_t(\mathbf{s})](\boldsymbol{\theta}) = (g_{\boldsymbol{\theta}}(\mathbf{X}_t, \mathbf{s}(\boldsymbol{\theta})), s_1(\boldsymbol{\theta}), \ldots, s_{q-1}(\boldsymbol{\theta}))^T, \qquad t \in \mathbb{Z}. \tag{5.30}$$

We can regard $\hat{\mathbf{h}}_t(\boldsymbol{\theta}) = (\hat{h}_t(\boldsymbol{\theta}), \ldots, \hat{h}_{t-q+1}(\boldsymbol{\theta}))^T$ as an "estimate" of the squared volatility vector $\boldsymbol{\sigma}_t^2$ under the parameter hypothesis $\boldsymbol{\theta}$, which is based on the data $X_{-p+1}, \ldots, X_t$. Also observe that $\hat{\mathbf{h}}_t(\boldsymbol{\theta}_0) = \hat{\boldsymbol{\sigma}}_t^2$ for all $t \in \mathbb{N}$.

For establishing the consistency of the QMLE it is essential that one can approximate $(\hat{h}_t)_{t\in\mathbb{N}}$ by a stationary ergodic sequence $(h_t)_{t\in\mathbb{N}}$ such that the error $\hat{h}_t - h_t$ converges to zero sufficiently fast as $t \to \infty$ and such that $(h_t(\boldsymbol{\theta})) = (\sigma_t^2)$ a.s. if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$; see Theorem 5.3.1. In particular, these requirements on $(h_t)_{t\in\mathbb{N}}$ comprehend the invertibility of the time series (X_t) because $\hat{h}_t(\boldsymbol{\theta}_0) = \hat{\sigma}_t^2$. We mention that invertibility is a common assumption in the classical theory of parameter estimation in ARMA time series. Only recently certain aspects of estimation in non–invertible linear time series have been studied, see e.g. Davis and Dunsmuir [34] or Breidt et al. [26].

For finding a candidate $(\mathbf{h}_t)_{t\in\mathbb{N}}$ for such an approaching sequence, first observe that $(\hat{\mathbf{h}}_t)_{t\in\mathbb{N}}$ is a solution of the SRE

$$\mathbf{s}_{t+1} = \Phi_t(\mathbf{s}_t), \qquad t \in \mathbb{N}, \tag{5.31}$$

on $\mathbb{C}(K,[0,\infty)^q)$ (provided certain regularity assumptions on $g_{\boldsymbol{\theta}}$ are fulfilled). Theorem 2.6.1 at hand, it is clear that if Φ_0 or one of its r–fold iterates is a contraction on average, then the unique stationary ergodic solution of the SRE (5.31) with index set $\mathbb{Z}$ provides the desired sequence $(\mathbf{h}_t)$. We summarize our findings in a proposition.

Proposition 5.2.12. *Assume that model* (5.1) *admits a unique stationary ergodic solution* $((X_t, \sigma_t))$ *and that the map* $(\boldsymbol{\theta}, \mathbf{s}) \mapsto g_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{s})$ *is continuous for every* $\mathbf{x} \in \mathbb{R}^p$*, which implies that* (Φ_t) *is a stationary ergodic sequence of mappings* $\mathbb{C}(K,[0,\infty)^q) \to \mathbb{C}(K,[0,\infty)^q)$*. We suppose the following conditions hold:*

S.1 $\mathbb{E}(\log^+ \|\Phi_0(\varsigma_0^2)\|_K) < \infty$.

S.2 $\mathbb{E}[\log^+ \Lambda(\Phi_0)] < \infty$ *and there exists an integer* $r \geq 1$ *such that* $\mathbb{E}[\log \Lambda(\Phi_0^{(r)})] < 0$.

Then the SRE (5.31) (*with the index set* $\mathbb{N}$ *replaced by* $\mathbb{Z}$) *has a unique stationary solution* $(\mathbf{h}_t)$*, which is ergodic. For every* $t \in \mathbb{Z}$*, the random elements* $\mathbf{h}_t$ *are* $\mathcal{F}_{t-1}$*–measurable and* $\mathbf{h}_t(\boldsymbol{\theta}_0) = \boldsymbol{\sigma}_t^2$ a.s. *Moreover,*

$$\|\hat{\mathbf{h}}_t - \mathbf{h}_t\|_K \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty. \tag{5.32}$$

Proof. The existence and uniqueness of a stationary ergodic solution to (5.31) and the limit relation (5.32) are a direct consequence of Theorem 2.6.1. It remains to give arguments for the other assertions. The backward iterates associated with (5.31) are given by

$$\mathbf{h}_{t,m} = \begin{cases} \varsigma_0^2, & m = 0, \\ \Phi_{t-1} \circ \cdots \circ \Phi_{t-m}(\varsigma_0^2), & m \geq 1. \end{cases} \tag{5.33}$$

This reveals that they are of form $\mathbf{h}_{t,m} = f_m(X_{t-1}, X_{t-2}, \ldots)$ for certain measurable maps f_m. Since $\mathbf{h}_{t,m} \xrightarrow{\text{a.s.}} \mathbf{h}_t$ as $t \to \infty$, an application of Corollary 2.1.3 shows that $\mathbf{h}_t = f(X_{t-1}, X_{t-2}, \ldots)$ a.s., where f is measurable. From this we conclude that $\mathbf{h}_t$ is $\mathcal{F}_{t-1}$–measurable for every t. The relation $\mathbf{h}_t(\boldsymbol{\theta}_0) = \boldsymbol{\sigma}_t^2$ a.s. follows from

$$\mathbf{h}_{t,m}(\boldsymbol{\theta}_0) - \boldsymbol{\sigma}_t^2 = \boldsymbol{\sigma}_{t,m}^2 - \boldsymbol{\sigma}_t^2 \stackrel{\text{d}}{=} \hat{\boldsymbol{\sigma}}_m^2 - \boldsymbol{\sigma}_m^2$$

together with $|\hat{\boldsymbol{\sigma}}_m^2 - \boldsymbol{\sigma}_m^2| \xrightarrow{\text{e.a.s.}} 0$, as shown in Proposition 5.2.6. □

5.3 Consistency of the QMLE

Suppose we observe data $X_{-p+1}, \ldots, X_0, X_1, \ldots, X_n$ generated by the model (5.1) with $\boldsymbol{\theta}_0$ as the true parameter; here we have to emphasize that by a shift of the index we can always assume that the data $X_{-p+1}, \ldots, X_0$ are available to us. By this convention, $\hat{h}_1$ is then well–defined. We set

$$\hat{L}_n(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^{n} \left(\frac{X_t^2}{\hat{h}_t(\boldsymbol{\theta})} + \log \hat{h}_t(\boldsymbol{\theta}) \right), \tag{5.34}$$

which can be regarded as an approximate conditional Gaussian likelihood up to some omitted constant, see Section 5.1. Note that $\hat{h}_t$ is the first coordinate of the random vector function $\hat{\mathbf{h}}_t = (\hat{h}_t, \ldots, \hat{h}_{t-q+1})^T$ defined in (5.29). The function $\hat{\mathbf{h}}_t(\boldsymbol{\theta})$ serves as an estimate of the vector $\boldsymbol{\sigma}_t^2 = (\sigma_t^2, \ldots, \sigma_{t-q+1}^2)^T$ under the parameter hypothesis $\boldsymbol{\theta}$, see Section 5.2.3. The (Gaussian) QMLE $\hat{\boldsymbol{\theta}}_n$ maximizes $\hat{L}_n$ on K, where K is an appropriately chosen subset of the parameter space Θ, i.e.,

$$\hat{\boldsymbol{\theta}}_n = \text{argmax}_{\boldsymbol{\theta} \in K} \hat{L}_n(\boldsymbol{\theta}) \,. \tag{5.35}$$

As a remark we mention that $\hat{\boldsymbol{\theta}}_n$ can be chosen to be measurable; this follows from a selection theorem going back to von Neumann [130], see e.g. Parthasarathy [111], Section 8.

Assume that the conditions of Proposition 5.2.12 are satisfied. These imply $||\hat{\mathbf{h}}_t - \mathbf{h}_t||_K \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$, where $(\mathbf{h}_t)$ is the unique stationary solution of the SRE $\mathbf{s}_{t+1} = \Phi_t(\mathbf{s}_t)$, $t \in \mathbb{Z}$. Note also that this is equivalent to

$$||\hat{h}_t - h_t||_K \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty.$$

We then define

$$L_n(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^{n} \left(\frac{X_t^2}{h_t(\boldsymbol{\theta})} + \log h_t(\boldsymbol{\theta}) \right) \tag{5.36}$$

together with

$$\tilde{\boldsymbol{\theta}}_n = \mathrm{argmax}_{\boldsymbol{\theta}\in K} L_n(\boldsymbol{\theta}). \tag{5.37}$$

From a theoretical point of view, it is more convenient to work with (L_n) because $(X_t^2/h_t + \log h_t)$ is stationary ergodic whereas $(X_t^2/\hat{h}_t + \log \hat{h}_t)_{t\in\mathbb{N}}$ is not. In what follows, we give a set of conditions which imply the strong consistency of $\hat{\boldsymbol{\theta}}_n$ (and $\tilde{\boldsymbol{\theta}}_n$):

C.1 Model (5.1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ admits a unique stationary ergodic solution $((X_t, \sigma_t))$ with $\mathbb{E}(\log^+ \sigma_0^2) < \infty$.

C.2 The conditions of Proposition 5.2.12 are fulfilled for a compact set $K \subset \Theta$ with $\boldsymbol{\theta}_0 \in K$.

C.3 The class of functions $\{g_{\boldsymbol{\theta}} \,|\, \boldsymbol{\theta} \in K\}$ is uniformly bounded from below, i.e., there exists a constant $\underline{g} > 0$ such that $g_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{s}) \geq \underline{g}$ for all $(\mathbf{x}, \mathbf{s}) \in \mathbb{R}^p \times [0, \infty)^q$ and $\boldsymbol{\theta} \in K$.

C.4 The following identifiability condition holds true on K: for all $\boldsymbol{\theta} \in K$,

$$h_0(\boldsymbol{\theta}) \equiv \sigma_0^2 \text{ a.s. if and only if } \boldsymbol{\theta} = \boldsymbol{\theta}_0.$$

C.5 The random elements σ_0^2/h_0 and $\log h_0$ have a finite expected norm:

$$\mathbb{E}\left\|\frac{\sigma_0^2}{h_0}\right\|_K < \infty \qquad \text{and} \qquad \mathbb{E}\|\log h_0\|_K < \infty.$$

These conditions are similar to those of Jeantheau [70].

Theorem 5.3.1. *Under the conditions* C.1 – C.4 *the* QMLE $\hat{\boldsymbol{\theta}}_n$ *is strongly consistent, i.e.,*

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0, \qquad n \to \infty.$$

Proof. In the first part we give the proof under the additional condition C.5 which allows one to apply Theorem 2.2.1 to L_n/n. In the second part we will indicate that C.5 is not needed. We have chosen to give two different proofs since the use of the uniform strong law of large numbers is intuitively more appealing than the proof without C.5.

Part 1. Assume C.1 – C.5. First we show that $\hat{L}_n/n \xrightarrow{\text{a.s.}} L$ in $\mathbb{C}(K)$ as $n \to \infty$, where

$$L(\boldsymbol{\theta}) = -\frac{1}{2}\mathbb{E}\left(\frac{\sigma_0^2}{h_0(\boldsymbol{\theta})} + \log h_0(\boldsymbol{\theta})\right), \quad \boldsymbol{\theta} \in K.$$

Secondly, we need to prove that L is uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. In the third step one shows that the a.s. uniform convergence of $\hat{L}_n/n$ towards L together with the fact that the limit L has a unique maximum implies strong consistency.

(i) We first establish $L_n/n \xrightarrow{\text{a.s.}} L$ in $\mathbb{C}(K)$ by an application of Theorem 2.2.1. Proposition 5.2.12 shows that the sequence $(\ell_t) = -2^{-1}(X_t^2/h_t + \log h_t)$ of random elements with values in $\mathbb{C}(K)$ is of the form $(\ell_t) = (f(X_t, X_{t-1}, \ldots))$, where f is measurable, and hence stationary ergodic (Proposition 2.1.1). Since $X_0 = \sigma_0 Z_0$ with Z_0 independent of σ_0 and h_0 and $\mathbb{E}Z_0^2 = 1$, assumption C.5 implies $\mathbb{E}\|X_0^2/h_0\|_K = \mathbb{E}\|\sigma_0^2/h_0\|_K < \infty$. Altogether, $\mathbb{E}\|\ell_0\|_K < \infty$, so that $L_n/n \xrightarrow{\text{a.s.}} L$ by Theorem 2.2.1. The property $\hat{L}_n/n \xrightarrow{\text{a.s.}} L$ follows if we can demonstrate $\|\hat{L}_n - L_n\|_K/n \xrightarrow{\text{a.s.}} 0$. Since $h_t, \hat{h}_t \geq \underline{g} > 0$, an application of the mean value theorem leads to $\|(\hat{h}_t)^{-1} - (h_t)^{-1}\|_K \leq \underline{g}^{-2}\|\hat{h}_t - h_t\|_K$ and $\|\log \hat{h}_t - \log h_t\|_K \leq \underline{g}^{-1}\|\hat{h}_t - h_t\|_K$. Thus there exists $c > 0$ with

$$\|\hat{L}_n - L_n\|_K \leq c\sum_{t=1}^{n}(1 + X_t^2)\|\hat{h}_t - h_t\|_K \leq c\sum_{t=1}^{\infty}(1 + X_t^2)\|\hat{h}_t - h_t\|_K.$$

By Proposition 5.2.12, $\|\hat{h}_t - h_t\|_K \xrightarrow{\text{e.a.s.}} 0$ and by condition C.1 together with $\mathbb{E}Z_0^2 = 1$ and Lemma 2.5.3 we have $\mathbb{E}[\log^+(1 + X_0^2)] < \infty$. An application of Proposition 2.5.1 demonstrates $\sum_{t=1}^{\infty}(1 + X_t^2)\|\hat{h}_t - h_t\|_K < \infty$ a.s. Hence $\|\hat{L}_n - L_n\|_K/n \xrightarrow{\text{a.s.}} 0$ and $\hat{L}_n/n \xrightarrow{\text{a.s.}} L$, as claimed.

(ii) For the uniqueness of the maximum of L on K we need to prove that $L(\boldsymbol{\theta}) < L(\boldsymbol{\theta}_0)$ for all $\boldsymbol{\theta} \in K\backslash\{\boldsymbol{\theta}_0\}$. Since $\mathbb{E}(\log \sigma_0^2)$ is finite and does not depend on the parameter $\boldsymbol{\theta}$, we can equivalently demonstrate that the function

$$Q(\boldsymbol{\theta}) = \mathbb{E}\left(\log\frac{\sigma_0^2}{h_0(\boldsymbol{\theta})} - \frac{X_0^2}{h_0(\boldsymbol{\theta})}\right) = \mathbb{E}\left(\log\frac{\sigma_0^2}{h_0(\boldsymbol{\theta})} - \frac{\sigma_0^2}{h_0(\boldsymbol{\theta})}\right), \quad \boldsymbol{\theta} \in K,$$

is uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. One can verify that $\log(x) - x \leq -1$ for all $x > 0$ with equality if and only if $x = 1$. Hence $Q(\boldsymbol{\theta}) \leq -1 = Q(\boldsymbol{\theta}_0)$ with equality if and only if $\sigma_0^2/h_0(\boldsymbol{\theta}) \equiv 1$ a.s. By C.4, $\sigma_0^2/h_0(\boldsymbol{\theta}) \equiv 1$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, which shows that Q and L are uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

(iii) Showing that (i) and (ii) imply strong consistency is accomplished by using standard arguments, which go back to Wald [131] and Le Cam [81]. Let $\epsilon > 0$ be arbitrary and suppose by contradiction that $\mathbb{P}(\limsup_{n\to\infty} |\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| > \epsilon) > 0$. Since the set $K' = K \cap \{\boldsymbol{\theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \geq \epsilon\}$ is compact and $\hat{L}_n/n \xrightarrow{\text{a.s.}} L$ in $\mathbb{C}(K)$, there is an event $D \subset \{\limsup_{n\to\infty} |\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| > \epsilon\}$ having positive probability and being such that for every $\omega \in D$ one can find a convergent subsequence $(\hat{\boldsymbol{\theta}}_{n_k}) \subset K'$ with $\lim_{k\to\infty} \hat{\boldsymbol{\theta}}_{n_k} = \boldsymbol{\theta}$ and $\hat{L}_{n_k}/n_k \to L$ in $\mathbb{C}(K)$; notice that (n_k) and $\boldsymbol{\theta} \in K'$ depend on the realization ω. On the other hand, by the definition of the QMLE, $L(\boldsymbol{\theta}) = \lim_{k\to\infty} \hat{L}_{n_k}(\hat{\boldsymbol{\theta}}_{n_k})/n_k \geq \lim_{k\to\infty} \hat{L}_{n_k}(\boldsymbol{\theta}_0)/n_k = L(\boldsymbol{\theta}_0)$ on D. Since $D \neq \emptyset$, there exists at least one point $\boldsymbol{\theta} \in K'$ with $L(\boldsymbol{\theta}) \geq L(\boldsymbol{\theta}_0)$, which contradicts the fact that L is uniquely maximized at $\boldsymbol{\theta}_0$. Consequently with probability one, $|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \leq \epsilon$ for all but finitely many n. Since $\epsilon > 0$ was arbitrary, we conclude $\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$ as $n \to \infty$.

Part 2. Without C.5 there is no longer uniform convergence of L_n/n towards L. The proof of strong consistency rests on an argument by Pfanzagl [112]. By virtue of Proposition 5.2.12 and C.3 the function

$$\boldsymbol{\theta} \mapsto \ell_t(\boldsymbol{\theta}) = -\frac{1}{2}\left(\frac{X_t^2}{h_t(\boldsymbol{\theta})} + \log h_t(\boldsymbol{\theta})\right)$$

is continuous on K with probability 1. Since for every *fixed* $\boldsymbol{\theta} \in K$ the sequence $(\ell_t(\boldsymbol{\theta}))$ is stationary ergodic, one has that $n^{-1}\sum_{t=1}^n \ell_t(\boldsymbol{\theta}) \xrightarrow{\text{a.s.}} L(\boldsymbol{\theta}) = \mathbb{E}[\ell_0(\boldsymbol{\theta})]$ as $n \to \infty$ by an application of the ergodic theorem for real random variables; note that if $\mathbb{E}X_0^2 = \infty$ the latter limit can take the value $-\infty$ at certain points $\boldsymbol{\theta}$, but $h_0(\boldsymbol{\theta}) \geq \underline{g} > 0$ guarantees $L(\boldsymbol{\theta}) < +\infty$ for all $\boldsymbol{\theta} \in K$. Therefore we can use exactly the same arguments as given in the proof of Lemma 3.11 of Pfanzagl [112] in order to show that the function L is upper–semi–continuous on K and $\limsup_{n\to\infty} \sup_{\boldsymbol{\theta}\in K'} L_n(\boldsymbol{\theta})/n \leq \sup_{\boldsymbol{\theta}\in K'} L(\boldsymbol{\theta})$ with probability 1 for any compact subset $K' \subseteq K$. Since C.1 – C.4 imply $\|\hat{L}_n - L_n\|_K/n \xrightarrow{\text{a.s.}} 0$, the inequality $\limsup_{n\to\infty} \sup_{\boldsymbol{\theta}\in K'} \hat{L}_n(\boldsymbol{\theta})/n \leq \sup_{\boldsymbol{\theta}\in K'} L(\boldsymbol{\theta})$ a.s. is valid also. Because an upper–semi–continuous function attains its maximum on compact sets, one can demonstrate $\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$, similarly to step (iii) of the proof of Part 1. □

5.4 Examples: Consistency

For the purpose of illustration we apply Theorem 5.3.1 to EGARCH and AGARCH(p, q). Our task will be to define the set K in an appropriate way and to verify the conditions C.1 – C.4.

5.4.1 EGARCH

Subsume the EGARCH parameters α,β,γ and δ into $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta)^T$ and denote by $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \gamma_0, \delta_0)^T$ the true parameter vector. As discussed in Example 5.2.8, we suppose $0 \leq \beta_0 < 1$, $\delta_0 \geq |\gamma_0|$. The parameter space is of form $\Theta = \mathbb{R} \times [0, 1) \times D_E$, where

$$D_E = \{(\gamma, \delta)^T \in \mathbb{R}^2 \,|\, \gamma \in \mathbb{R},\, \delta \geq |\gamma|\,\}.$$

The compact set $K \subset \Theta$ will be defined below. The restriction $0 \leq \beta_0 < 1$ guarantees that the SRE (5.11) has a unique stationary ergodic solution $(\log \sigma_t^2)$. From the almost sure representation (5.12) of this solution, one recognizes $\mathbb{E}(\log \sigma_0^2) < \infty$, which establishes C.1.

Rather than checking condition C.2 of Theorem 5.3.1, we *directly* verify its consequences which were used in the proof of the latter theorem, namely the fact that a stationary ergodic sequence (h_t) of random elements with values in $\mathbb{C}(K,[0,\infty))$ can be defined such that

$$h_t \text{ is } \mathcal{F}_{t-1}\text{-measurable and } h_t(\boldsymbol{\theta}_0) = \sigma_t^2 \text{ a.s. for every } t, \tag{5.38}$$

$$X_t^2 \|\hat{h}_t^{-1} - h_t^{-1}\|_K \xrightarrow{\text{e.a.s.}} 0 \text{ and } \|\log \hat{h}_t - \log h_t\|_K \xrightarrow{\text{e.a.s.}} 0 \text{ as } t \to \infty. \tag{5.39}$$

To this end, we consider the SRE

$$\log s_{t+1} = \Phi_t(\log s_t), \qquad t \in \mathbb{Z}, \tag{5.40}$$

where

$$[\Phi_t(s)](\boldsymbol{\theta}) = \alpha + \beta s(\boldsymbol{\theta}) + (\gamma X_t + \delta|X_t|)\exp(-s(\boldsymbol{\theta})/2), \qquad \boldsymbol{\theta} \in K.$$

Observing that $(\gamma X_t + \delta|X_t|)\exp(-s(\boldsymbol{\theta})/2) \geq 0$, we find that for any constant function $s = \log\varsigma_0^2$ and $\epsilon > 0$ the forward (and backward) iterates associated with (5.40) fulfill

$$\begin{aligned}\log \hat{h}_t(\boldsymbol{\theta}) = [\Phi_{t-1}^{(t)}(\log\varsigma_0^2)](\boldsymbol{\theta}) &\geq \alpha(1+\beta+\cdots+\beta^{t-1}) + \beta^t \log\varsigma_0^2 \\ &\geq \alpha(1-\beta)^{-1} - \epsilon\end{aligned}$$

for large enough t. Therefore we may suppose without loss of generality that the SRE (5.40) lives on the subset $\mathbb{C}(K,[m-\epsilon,\infty))$ with $m = \inf_{\boldsymbol{\theta}\in K} \alpha(1-\beta)^{-1}$. Note also that $\mathbb{C}(K,[m-\epsilon,\infty))$ is a complete and separable metric space. By a comparison with the derivation of (5.24), one recognizes that

$$\Lambda(\Phi_0) = \sup_{\boldsymbol{\theta}\in K} \lambda_\epsilon(\boldsymbol{\theta}) = \|\lambda_\epsilon\|_K,$$

where

$$\lambda_\epsilon(\boldsymbol{\theta}) = \max\big(\beta, 2^{-1}\exp(-(m-\epsilon)/2)(\gamma X_0 + \delta|X_0|) - \beta\big). \tag{5.41}$$

Since $\mathbb{E}(\log\|\lambda_\epsilon\|_K) \to \mathbb{E}(\log\|\lambda_0\|_K)$ as $\epsilon \downarrow 0$, the condition $\mathbb{E}(\log\|\lambda_0\|_K) < 0$ implies $\mathbb{E}(\log\|\lambda_\epsilon\|_K) < 0$ for ϵ small enough. Hence, if we assume K is chosen such that $\mathbb{E}(\log\|\lambda_0\|_K) < 0$, then the sequence $(\Lambda(\Phi_t))$ obeys the conditions of Theorem 2.6.1, and thus the SRE (5.40) (on $\mathbb{C}(K,[m-\epsilon,\infty))$, where $\epsilon > 0$ small) admits a unique stationary solution $(\log h_t)$, which is ergodic. Moreover $\log h_t$ is $\mathcal{F}_{t-1}$–measurable and the property $h_t(\boldsymbol{\theta}) = h_t(\boldsymbol{\theta}_0)$ a.s. follows from similar arguments as in the proof of Proposition 5.2.12. Thus (5.38) is established. As regards the verification of (5.39), note that $\|\log\hat{h}_t - \log h_t\|_K \xrightarrow{\text{e.a.s.}} 0$ by Theorem 2.6.1. The mean value theorem applied to the function $e^{-x/2}$ together with the facts that $\log\hat{h}_t \geq m - \epsilon$ for t large and $\log h_t \geq m$ for all t yields a constant $c > 0$ with

$$\begin{aligned}X_t^2\,\|(\hat{h}_t)^{-1} - (h_t)^{-1}\|_K &= X_t^2\,\|\exp(-(\log\hat{h}_t)/2) - \exp(-(\log h_t)/2)\|_K \\ &\leq cX_t^2\,\|\log\hat{h}_t - \log h_t\|_K \xrightarrow{\text{e.a.s.}} 0.\end{aligned}$$

The latter limit relation is a consequence of $\mathbb{E}(\log^+ X_0^2) < \infty$, implied by $\mathbb{E}(\log \sigma_0^2) < \infty$ and $\mathbb{E}Z_0^2 = 1$, and an application of Proposition 2.5.1. This completes the demonstration of the limit relations (5.39).

Since condition C.3 is automatically fulfilled for the EGARCH process, we are left with the verification of the identifiability condition C.4. Before we start, we need to impose the mild technical assumptions that $(\gamma_0, \delta_0) \neq (0,0)$ and that the distribution of Z_0 is not concentrated in two points. Note that the case $\gamma_0 = \delta_0 = 0$ would lead to an identifiability problem because then $\log \sigma_t^2 = \alpha_0(1-\beta_0)^{-1}$ by representation (5.12), which implies that there are infinitely many parameters leading to the identical model. Observe that $h_0(\boldsymbol{\theta}) = \sigma_0^2$ a.s. is equivalent to $\log h_t(\boldsymbol{\theta}) = \log h_t(\boldsymbol{\theta}_0)$ a.s. for all $t \in \mathbb{Z}$ because of the stationarity of $(\log h_t - \sigma_t^2)$ and the property $\log h_0(\boldsymbol{\theta}_0) = \log \sigma_0^2$. We now show the nontrivial implication $\log h_t(\boldsymbol{\theta}) = \log h_t(\boldsymbol{\theta}_0) \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0$. Replacing $\log h_t$ by $\Phi_{t-1}(\log h_{t-1})$ in the identity $\log h_t(\boldsymbol{\theta}) = \log h_t(\boldsymbol{\theta}_0)$ and accounting for $h_{t-1}(\boldsymbol{\theta}) = \sigma_{t-1}^2$, we obtain

$$(\alpha - \alpha_0) + (\beta - \beta_0)\log \sigma_{t-1}^2 + \{\, (\gamma - \gamma_0) Z_{t-1} + (\delta - \delta_0)|Z_{t-1}| \,\} = 0 \quad \text{a.s.}$$

If $\beta \neq \beta_0$, the random variable $\log \sigma_{t-1}^2$ would at the same time be a measurable function of Z_{t-1} and independent of Z_{t-1}. By Lemma 5.4.2 below this implies that $\log \sigma_{t-1}^2$ is deterministic. However, taking the variance in (3.32) gives $\operatorname{Var}(\log \sigma_{t-1}^2) = \sum_{k=0}^{\infty} \beta_0^{2k} \operatorname{Var}(\gamma_0 Z_0 + \delta_0 |Z_0|) > 0$ since the facts that $(\gamma_0, \delta_0) \neq (0,0)$ and that the distribution of Z_0 is not concentrated in two points imply $\operatorname{Var}(\gamma_0 Z_0 + \delta_0 |Z_0|) > 0$. To avoid this contradiction, necessarily $\beta = \beta_0$. Furthermore, if $(\alpha - \alpha_0, \gamma - \gamma_0, \delta - \delta_0) \neq (0,0,0)$, there are three distinct cases concerning the roots of the function $f(z) = (\alpha_0 - \alpha_0) + (\gamma - \gamma_0) z + (\delta - \delta_0)|z|$:

1. f has less than or equal to 2 roots.
2. $f \equiv 0$ on $[0, \infty)$ and $f \neq 0$ on $(-\infty, 0)$.
3. $f \neq 0$ on $(0, \infty)$ and $f \equiv 0$ on $(-\infty, 0]$.

Using this observation and the facts that $\mathbb{E}Z_0 = 0$, $\mathbb{E}Z_0^2 = 1$ and that the distribution of Z_0 is not concentrated in two points, we conclude that $f(Z_{t-1}) = 0$ a.s. if and only if $(\alpha, \beta, \gamma) = (\alpha_0, \gamma_0, \delta_0)$. Thus $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, which concludes the verification of C.4. Therefore the strong consistency of the QMLE in EGARCH can be established by an application of Theorem 5.3.1.

Theorem 5.4.1. *Let (X_t) be a stationary* EGARCH *process with parameters $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \gamma_0, \delta_0)^T$ such that $(\gamma_0, \delta_0) \neq (0,0)$. Suppose the distribution of Z_0 is not concentrated in two points. Let $K \subset \mathbb{R} \times [0,1) \times D_E$ be a compact set with $\boldsymbol{\theta}_0 \in K$ and such that*

$$\mathbb{E}(\log \|\lambda_0\|_K) < 0,$$

where λ_0 is given by (5.41) *with $m = \inf_{\boldsymbol{\theta} \in K} \alpha(1-\beta)^{-1}$ and $\epsilon = 0$. Then the* QMLE *$\hat{\boldsymbol{\theta}}_n$ is strongly consistent.*

Eventually we provide the previously announced auxiliary lemma.

Lemma 5.4.2. *Let the real random variable U be at the same time measurable with respect to a σ–algebra $\mathcal{A}$ and independent of $\mathcal{A}$. Then U is necessarily a constant.*

Proof. The proof is by contradiction. If U is not a constant, there exists $a \in \mathbb{R}$ such that $0 < \mathbb{P}(U \le a) < 1$. Then since the event $\{U \le a\} \in \mathcal{A}$ is independent of itself,

$$\mathbb{P}(U \le a) = \mathbb{P}(\{U \le a\}, \{U \le a\}) = (\mathbb{P}(U \le a))^2.$$

This contradicts $0 < \mathbb{P}(U \le a) < 1$ and concludes the proof. □

5.4.2 AGARCH(p,q)

Set $\boldsymbol{\theta} = (\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \gamma)^T$ and write $\boldsymbol{\theta}_0 = (\alpha_0^\circ, \ldots, \alpha_p^\circ, \beta_1^\circ, \ldots, \beta_q^\circ, \gamma^\circ)^T$ for the true parameter vector of AGARCH(p,q). We suppose that $\boldsymbol{\theta}_0$ admits a unique stationary ergodic solution $((X_t, \sigma_t))$ to the AGARCH(p,q) equations (3.8)–(3.9), which is equivalent to a negative top Lyapunov exponent of the associated matrix sequence $(\mathbf{A}_t)$ given by (5.14) with $\alpha_i = \alpha_i^\circ$, $\beta_j = \beta_j^\circ$ and $\gamma = \gamma^\circ$; see Example 5.2.5 or Theorem 3.3.1. Moreover, suppose $\alpha_i^\circ > 0$ for some $i = 1, \ldots, p$ because otherwise the constant sequence $\sigma_t^2 = \alpha_0^\circ(1 - \sum_{j=1}^q \beta_j^\circ)^{-1}$ is the unique stationary solution of (5.13), which would imply that one cannot discriminate between α_0° and the β_j°'s (non–identifiability). Another necessary restriction is $(\alpha_p^\circ, \beta_q^\circ) \neq (0,0)$. Let K be a compact subset of $(0,\infty) \times [0,\infty)^p \times B \times [-1,1]$ containing the true parameter $\boldsymbol{\theta}_0$, where

$$B = \Big\{ (\beta_1, \ldots, \beta_q)^T \in [0,1)^q \,\Big|\, \sum_{j=1}^q \beta_j < 1 \Big\}. \tag{5.42}$$

We now verify C.1 – C.4.

In Example 5.2.5 we have shown the existence of a $\eta > 0$ with $\mathbb{E}\sigma_0^{2\eta} < \infty$. Thus $\mathbb{E}(\log^+ \sigma_0^2) < \infty$ and C.1 is valid.

As regards C.2, note that the random maps (5.30) are of form

$$[\Phi_t(\mathbf{s})](\boldsymbol{\theta}) = \Big(\alpha_0 + \sum_{i=1}^p \alpha_i (|X_{t+1-i}| - \gamma X_{t+1-i})^2 \Big) \mathbf{e}_1 + \mathbf{C}(\boldsymbol{\theta}) \mathbf{s}(\boldsymbol{\theta}),$$

for $\mathbf{s} \in \mathbb{C}(K, [0,\infty)^q)$, where

$$\mathbf{e}_1 = (1, 0 \ldots, 0)^T \in \mathbb{R}^q,$$

$$\mathbf{C}(\boldsymbol{\theta}) = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \cdots & \beta_{q-1} & \beta_q \\ 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 1 & 0 \end{pmatrix} \in \mathbb{R}^{q \times q}.$$

Compare with Example 5.2.11. Note that $\Lambda(\Phi_0^{(r)}) = \Lambda(\mathbf{C}^r)$, $r \in \mathbb{N}$, where $\mathbf{C}^r$ has to be understood as the map $\mathbf{s} \mapsto \mathbf{C}^r\mathbf{s}$ on $\mathbb{C}(K, [0, \infty)^q)$. We recognize $\bar{\beta} := \sup_{\boldsymbol{\theta} \in K} \left(\sum_{j=1}^q \beta_j\right) < 1$ since K is compact. A pointwise application of the inequality (5.27) yields the bound

$$|(\mathbf{C}^r\mathbf{s})(\boldsymbol{\theta})| \le \sqrt{q} \left(\sum_{j=1}^q \beta_j\right)^{[r/q]} |\mathbf{s}(\boldsymbol{\theta})|, \qquad \boldsymbol{\theta} \in K,$$

for any $\mathbf{s} \in \mathbb{C}(K, [0, \infty)^q)$. Taking the supremum on both sides of the latter bound one obtains

$$\|\mathbf{C}^r\mathbf{s}\|_K \le \sqrt{q}\, \bar{\beta}^{[r/q]} \|\mathbf{s}\|_K, \qquad r \ge 0,$$

showing that $\Lambda(\mathbf{C}^r) \le \sqrt{q}\, \bar{\beta}^{[r/q]} \to 0$ as $r \to \infty$. Therefore condition S.2 of Proposition 5.2.12 has been verified. It is standard to check S.1. Hence C.2 holds true and $\mathbf{h}_t = (h_t, \ldots, h_{t-q+1})^T$ is properly defined by virtue of Proposition 5.2.12.

C.3 being obviously fulfilled, we turn to the identifiability condition C.4. We split our arguments into a series of lemmas. First we derive an almost sure representation of h_t, similarly to Berkes et al. [8].

Lemma 5.4.3. *The following almost sure representation for h_t is valid:*

$$h_t(\boldsymbol{\theta}) = \pi_0(\boldsymbol{\theta}) + \sum_{\ell=1}^{\infty} \pi_\ell(\boldsymbol{\theta})(|X_{t-\ell}| - \gamma X_{t-\ell})^2, \qquad \boldsymbol{\theta} \in K, \tag{5.43}$$

with the sequence $(\pi_\ell(\boldsymbol{\theta}))_{\ell \in \mathbb{N}}$ given by

$$\pi_0(\boldsymbol{\theta}) = \frac{\alpha_0}{\beta_{\boldsymbol{\theta}}(1)} = \alpha_0 \left(1 - \sum_{j=1}^q \beta_j\right)^{-1} \quad \text{and} \quad \sum_{\ell=1}^{\infty} \pi_\ell(\boldsymbol{\theta}) z^\ell = \frac{\alpha_{\boldsymbol{\theta}}(z)}{\beta_{\boldsymbol{\theta}}(z)}, \qquad |z| \le 1, \tag{5.44}$$

where $\alpha_{\boldsymbol{\theta}}(z) = \sum_{i=1}^p \alpha_i z^i$ and $\beta_{\boldsymbol{\theta}}(z) = 1 - \sum_{j=1}^q \beta_j z^j$.

Proof. The proof rests on the observation that $(h_t(\boldsymbol{\theta}))$ obeys an ARMA(p,q) equation, i.e.,

$$h_t(\boldsymbol{\theta}) = \alpha_0 + \sum_{i=1}^{p} \alpha_i(|X_{t-i}| - \gamma X_{t-i})^2 + \sum_{j=1}^{q} \beta_j h_{t-j}(\boldsymbol{\theta}), \qquad t \in \mathbb{Z}, \quad (5.45)$$

or shorter, in backshift operator notation,

$$\beta_{\boldsymbol{\theta}}(B) h_t(\boldsymbol{\theta}) = \alpha_0 + \alpha_{\boldsymbol{\theta}}(B)(|X_t| - \gamma X_t)^2, \qquad t \in \mathbb{Z}. \quad (5.46)$$

Since $\bar{\beta} < 1$, the roots of the polynomial $\beta_{\boldsymbol{\theta}}(z)$ lie outside the unit disc; indeed, if $|z| < \bar{\beta}^{-1/q}$ then $|\beta_{\boldsymbol{\theta}}(z)| \geq 1 - \sum_{j=1}^{q} \beta_j |z|^j > 1 - \bar{\beta}(\bar{\beta}^{-1/q})^q = 0$. This suggests the a.s. representation (5.43), as can be seen from a comparison with Section 3.2 or Chapter 3 of Brockwell and Davis [29]. To prove this a.s. representation we can however not directly apply Proposition 3.1.2 in Brockwell and Davis [29] because the "innovations" $(|X_t| - \gamma X_t)^2$ may have an infinite moment and because (5.45) depends on the parameter $\boldsymbol{\theta}$. One possible way to validate the a.s. representation is to show that the right–hand side of (5.43) is well–defined, continuous on K a.s., and that it obeys (5.46) regarded as a difference equation on $\mathbb{C}(K)$. Then, since the SRE for $(\mathbf{h}_t)$ admits a *unique* stationary ergodic solution $(\mathbf{h}_t)$, as follows from Proposition 5.2.12, the right–hand side of (5.43) must coincide with h_t. To show the three assertions mentioned before, first note that there are $0 < \lambda < 1$ and $c > 0$ with $|\pi_\ell(\boldsymbol{\theta})| \leq c\lambda^\ell$ for all $\ell \geq 1$ and $\boldsymbol{\theta} \in K$ (apply the Cauchy inequalities to the complex function $1/\beta_{\boldsymbol{\theta}}(z)$). Thus $\pi_\ell \to 0$ in $\mathbb{C}(K)$ with geometric rate. Since π_ℓ is continuous on K and $\mathbb{E}|\,|X_0| - \gamma^\circ X_0\,|^{2\eta} < \infty$ for an $\eta > 0$ (see Example 5.2.5), the series $\sum_{\ell=1}^{\infty} \pi_\ell(\boldsymbol{\theta})(|X_{t-\ell}| - \gamma X_{t-\ell})^2$ converges absolutely a.s. in $\mathbb{C}(K)$ by virtue of Proposition 2.5.1. Hence (5.43) is continuous a.s. Eventually it is an elementary exercise to prove that the a.s. representation for (h_t) obeys (5.46). Indeed, plugging (5.43) into the right–hand side of (5.45) and bearing in mind that the series $\sum_{\ell=1}^{\infty} \pi_\ell(\boldsymbol{\theta})(|X_{t-\ell}| - \gamma X_{t-\ell})^2$ converges absolutely for all t, we obtain

$$\begin{aligned}&\alpha_0 + \alpha_{\boldsymbol{\theta}}(B)(|X_t| - \gamma X_t)^2 + (1 - \beta_{\boldsymbol{\theta}}(B)) h_t(\boldsymbol{\theta}) \\ &\qquad = \alpha_0 + (1 - \beta_{\boldsymbol{\theta}}(1))\alpha_0(\beta_{\boldsymbol{\theta}}(1))^{-1} + (\alpha_{\boldsymbol{\theta}}(B) + \chi_{\boldsymbol{\theta}}(B))(|X_t| - \gamma X_t)^2,\end{aligned}$$

where $\chi_{\boldsymbol{\theta}}(z) = (1 - \beta_{\boldsymbol{\theta}}(z)) \sum_{\ell=1}^{\infty} \pi_\ell(\boldsymbol{\theta}) z^\ell$. By $\sum_{\ell=1}^{\infty} \pi_\ell(\boldsymbol{\theta}) z^\ell = \alpha_{\boldsymbol{\theta}}(z)/\beta_{\boldsymbol{\theta}}(z)$, one has $\alpha_{\boldsymbol{\theta}}(z) + \chi_{\boldsymbol{\theta}}(z) = \alpha_{\boldsymbol{\theta}}(z)/\beta_{\boldsymbol{\theta}}(z)$ and hence the right–hand side of the latter display coincides with (5.43). This completes the proof of the lemma. □

The next lemma is concerned with the identifiability of the parameter γ.

Lemma 5.4.4. *Suppose that the distribution of Z_0 is not concentrated in two points. Then for any $\boldsymbol{\theta} \in K$ the relation $h_0(\boldsymbol{\theta}) = h_0(\boldsymbol{\theta}_0)$* a.s. *implies $\gamma = \gamma^\circ$.*

Proof. Note that $h_0(\boldsymbol{\theta}) = h_0(\boldsymbol{\theta}_0)$ a.s. is equivalent to $h_k(\boldsymbol{\theta}) = h_k(\boldsymbol{\theta}_0)$ for any k, in particular $k = \max(p, q)$. We rewrite $h_k(\boldsymbol{\theta}) = h_k(\boldsymbol{\theta}_0)$ a.s. by

$$(\alpha_0^\circ - \alpha_0) + \sum_{i=1}^{k} Y_{k-i}\sigma_{k-i}^2 = 0 \quad \text{a.s.}, \tag{5.47}$$

where $Y_{k-i} = \alpha_i^\circ(|Z_{k-i}| - \gamma^\circ Z_{k-i})^2 - \alpha_i(|Z_{k-i}| - \gamma Z_{k-i})^2 + (\beta_i^\circ - \beta_i)$. Here we define $\alpha_i, \alpha_i^\circ = 0$ if $i \in (p, k]$ and $\beta_i, \beta_i^\circ = 0$ if $i \in [q, k)$. Introduce $k^* = \min(i \in [1, p] \mid \alpha_i^\circ > 0)$. Then by repeatedly expressing each term σ_{t-j}^2, $j = 1, \ldots, (k^* - 1)$, in the equation

$$\sigma_t^2 = \alpha_0^\circ + \sum_{i=k^*}^{p} \alpha_i^\circ(|X_{t-i}| - \gamma^\circ X_{t-i})^2 + \sum_{j=1}^{q} \beta_j^\circ \sigma_{t-j}^2$$

by past observations and past squared volatilities, one sees that σ_t^2 can be written as a function of

$$\{(|X_{t-k^*}| - \gamma^\circ X_{t-k^*})^2, (|X_{t-k^*-1}| - \gamma^\circ X_{t-k^*-1})^2, \ldots ; \sigma_{t-k^*}^2, \sigma_{t-k^*-1}^2, \ldots\},$$

and consequently σ_t^2 is $\mathcal{F}_{t-k^*}$–measurable. Relation (5.47) together with $\sigma_{k-1}^2 \geq \alpha_0^\circ > 0$ implies that Y_{k-1} is a function of $\sigma_0^2, \ldots, \sigma_{k-1}^2$ and consequently $\mathcal{F}_{k-k^*-1}$–measurable. Since Y_{k-1} is at the same time independent of $\mathcal{F}_{k-k^*-1}$, it must be degenerate (Lemma 5.4.2). With the identical arguments, $Y_{k-2}, \ldots, Y_{k-k^*}$ are degenerate. The degeneracy of Y_{k-k^*} means that

$$\alpha_{k^*}^\circ(|Z_{k-k^*}| - \gamma^\circ Z_{k-k^*})^2 - \alpha_{k^*}(|Z_{k-k^*}| - \gamma Z_{k-k^*})^2 = c$$

for a certain constant c. Note that with probability one on the sets $\{Z_{k-k^*} \geq 0\}$ and $\{Z_{k-k^*} < 0\}$,

$$\left(\alpha_{k^*}^\circ(1 - \gamma^\circ)^2 - \alpha_{k^*}(1 - \gamma)^2\right) Z_{k-k^*}^2 = c$$

and

$$\left(\alpha_{k^*}^\circ(1 + \gamma^\circ)^2 - \alpha_{k^*}(1 + \gamma)^2\right) Z_{k-k^*}^2 = c,$$

respectively. Because the distribution of Z_{k-k^*} is not concentrated in two points, these two equations can only be jointly fulfilled if $c = 0$. Since $\mathbb{E}Z_{k-k^*} = 0$ and $\mathbb{E}Z_{k-k^*}^2 = 1$, the distribution of Z_{k-k^*} has positive mass both on the negative and the positive real line, which implies

$$\alpha_{k^*}^\circ(1 - \gamma^\circ)^2 = \alpha_{k^*}(1 - \gamma)^2 \quad \text{and} \quad \alpha_{k^*}^\circ(1 + \gamma^\circ)^2 = \alpha_{k^*}(1 + \gamma)^2.$$

From these two equations together with $\alpha_{k^*}^\circ > 0$ we conclude $\alpha_{k^*} > 0$. Subtracting and adding the latter two equations yields $\gamma = \gamma^\circ \alpha_{k^*}^\circ / \alpha_{k^*}$ and $1 + \gamma^2 = (1 + (\gamma^\circ)^2)\alpha_{k^*}^\circ / \alpha_{k^*}$. Because of the constraint $|\gamma| \leq 1$ we can conclude $\gamma = \gamma^\circ$ (and $\alpha_{k^*} = \alpha_{k^*}^\circ$), which completes the proof. □

Eventually we establish the identifiability condition C.4.

Lemma 5.4.5. *Suppose that the distribution of* Z_0 *is not concentrated in two points and that the polynomials* $\alpha_{\boldsymbol{\theta}_0}$ *and* $\beta_{\boldsymbol{\theta}_0}$ *defined in Lemma 5.4.3 do not have any common roots. Then for any* $\boldsymbol{\theta} \in K$,

$$h_0(\boldsymbol{\theta}) = \sigma_0^2 \quad \textit{if and only if} \quad \boldsymbol{\theta} = \boldsymbol{\theta}_0.$$

Proof. We have shown in Lemma 5.4.4 that $h_0(\boldsymbol{\theta}) = \sigma_0^2$ implies $\gamma = \gamma^\circ$, so that in consideration of (5.43) the relation $h_0(\boldsymbol{\theta}) = h_0(\boldsymbol{\theta}_0)$ becomes

$$\pi_0(\boldsymbol{\theta}) - \pi_0(\boldsymbol{\theta}_0) + \sum_{\ell=1}^{\infty}(\pi_\ell(\boldsymbol{\theta}) - \pi_\ell(\boldsymbol{\theta}_0))(|X_{-\ell}| - \gamma^\circ X_{-\ell})^2 \equiv 0.$$

We first show $\pi_\ell(\boldsymbol{\theta}) - \pi_\ell(\boldsymbol{\theta}_0) = 0$ for all $\ell \in \mathbb{N}$ by contradiction. Denote by $\ell^* \geq 1$ the smallest integer $\ell \geq 1$ with $\delta_\ell := \pi_\ell(\boldsymbol{\theta}) - \pi_\ell(\boldsymbol{\theta}_0) \neq 0$. Since $\sigma^2_{-\ell^*} \geq \alpha_0^\circ > 0$, we have that

$$\begin{aligned}&(|Z_{-\ell^*}| - \gamma^\circ Z_{-\ell^*})^2 \\ &= \Big(\pi_0(\boldsymbol{\theta}_0) - \pi_0(\boldsymbol{\theta}) + \sum_{\ell=\ell^*+1}^{\infty}(\pi_\ell(\boldsymbol{\theta}_0) - \pi_\ell(\boldsymbol{\theta}))(|X_{-\ell}| - \gamma^\circ X_{-\ell})^2\Big)/(\delta_{\ell^*}\sigma^2_{-\ell^*})\end{aligned}$$

is at the same time $\mathcal{F}_{-\ell^*-1}$–measurable and independent of $\mathcal{F}_{-\ell^*-1}$, which by Lemma 5.4.2 is only possible if $(|Z_{-\ell^*}| - \gamma^\circ Z_{-\ell^*})^2$ is degenerate. However, from the assumption that the distribution of Z_0 is not concentrated in two points it follows that $(|Z_{-\ell^*}| - \gamma^\circ Z_{-\ell^*})^2$ cannot be degenerate, i.e. the desired contradiction. Using $\pi_\ell(\boldsymbol{\theta}) = \pi_\ell(\boldsymbol{\theta}_0)$ in (5.44), we conclude $\alpha_{\boldsymbol{\theta}}(z)/\beta_{\boldsymbol{\theta}}(z) = \alpha_{\boldsymbol{\theta}_0}(z)/\beta_{\boldsymbol{\theta}_0}(z)$. Write $\alpha_{\boldsymbol{\theta}}(z) = r(z)\alpha_{\boldsymbol{\theta}_0}(z)$ and $\beta_{\boldsymbol{\theta}}(z) = r(z)\beta_{\boldsymbol{\theta}_0}(z)$. The rational function $r(z)$ does not have any pole because otherwise $\alpha_{\boldsymbol{\theta}_0}(z)$ and $\beta_{\boldsymbol{\theta}_0}(z)$ would have a common root. Hence $r(z)$ is a polynomial. The degree of r is zero, because otherwise either $\alpha_{\boldsymbol{\theta}}(z)$ or $\beta_{\boldsymbol{\theta}}(z)$ would have degree strictly greater than p or q, respectively, since $(\alpha_p^\circ, \beta_q^\circ) \neq (0,0)$. Finally, $r \equiv 1$ because the constants in the polynomials $\beta_{\boldsymbol{\theta}}$ and $\beta_{\boldsymbol{\theta}_0}$ are 1. Hence $\alpha_{\boldsymbol{\theta}} = \alpha_{\boldsymbol{\theta}_0}$ and $\beta_{\boldsymbol{\theta}} = \beta_{\boldsymbol{\theta}_0}$, which gives $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and concludes the proof. □

Now an application of Theorem 5.3.1 yields the strong consistency of the QMLE in AGARCH(p,q). Mutatis mutandis the result is true also for GARCH(p,q), cf. Theorem 4.2.1. The necessary notational changes and modifications of the proofs are evident.

Theorem 5.4.6. *Let* (X_t) *be a stationary* AGARCH(p,q) *process with true parameters* $\boldsymbol{\theta}_0 = (\alpha_0^\circ, \ldots, \alpha_p^\circ, \beta_1^\circ, \ldots, \beta_q^\circ, \gamma^\circ)^T$ *such that*

1. $\alpha_i^\circ > 0$ *for some* $i > 0$ *and* $(\alpha_p^\circ, \beta_q^\circ) \neq (0,0)$.
2. *The polynomials* $\alpha^\circ(z) = \sum_{i=1}^p \alpha_i^\circ z^i$ *and* $\beta^\circ(z) = 1 - \sum_{j=1}^q \beta_j^\circ z^j$ *do not have any common roots.*

Suppose that the distribution of Z_0 *is not concentrated in two points and let* $K \subset (0,\infty) \times [0,\infty)^p \times B \times [-1,1]$ *be compact and contain* $\boldsymbol{\theta}_0$*, where* B *is given in* (5.42)*. Then the* QMLE $\hat{\boldsymbol{\theta}}_n$ *is strongly consistent.*

5.5 The First and Second Derivatives of $\mathbf{h}_t$ and $\hat{\mathbf{h}}_t$

For establishing the asymptotic normality of the QMLE, it is essential to understand the limit behavior of the sequences of functions $(\hat{h}_t')_{t\in\mathbb{N}}$ and $(\hat{h}_t'')_{t\in\mathbb{N}}$ and to study the differentiability properties of h_t; the symbol $'$ denotes differentiation with respect to $\boldsymbol{\theta}$. The interpretation of the arising problems in terms of SREs is fruitful once again. The results obtained here will prove useful in Section 5.6.

We recall from Section 5.2.3 that

$$\hat{\mathbf{h}}_{t+1} = \Phi_t(\hat{\mathbf{h}}_t), \qquad t \in \mathbb{N}, \tag{5.48}$$

where the random maps Φ_t on $\mathbb{C}(K,[0,\infty)^q)$ are defined in (5.30). Assume that the SRE associated with (Φ_t) is contractive, i.e., (Φ_t) obeys the conditions of Theorem 2.6.1 (or Proposition 5.2.12). The SREs for the first and second derivatives of $\hat{\mathbf{h}}_t$ are basically determined by (Φ_t). An important insight will be the facts that $(\hat{\mathbf{h}}_t')_{t\in\mathbb{N}}$ and $(\hat{\mathbf{h}}_t'')_{t\in\mathbb{N}}$ are solutions of contractive SREs "up to an exponentially fast decaying perturbance". By this we mean that

$$\hat{\mathbf{h}}_{t+1}' = \hat{\dot{\Phi}}_t(\hat{\mathbf{h}}_t') \quad \text{and} \quad \hat{\mathbf{h}}_{t+1}'' = \hat{\ddot{\Phi}}_t(\hat{\mathbf{h}}_t''), \qquad t \in \mathbb{N},$$

where the *nonstationary* sequences of maps $(\hat{\dot{\Phi}}_t)_{t\in\mathbb{N}}$ and $(\hat{\ddot{\Phi}}_t)_{t\in\mathbb{N}}$ can be approximated by stationary sequences $(\dot{\Phi}_t)_{t\in\mathbb{N}}$ and $(\ddot{\Phi}_t)_{t\in\mathbb{N}}$ in the sense of Theorem 2.6.4, and moreover the SREs

$$\mathbf{q}_{t+1} = \dot{\Phi}_t(\mathbf{q}_t) \quad \text{and} \quad \mathbf{q}_{t+1} = \ddot{\Phi}_t(\mathbf{q}_t), \qquad t \in \mathbb{Z}, \tag{5.49}$$

are contractive. These statements are valid under relatively weak additional assumptions in Proposition 5.2.12. Then it is evident from Theorem 2.6.1 that the SREs (5.49) have unique stationary ergodic solutions $(\mathbf{d}_t)$ and $(\tilde{\mathbf{d}}_t)$. Moreover Theorem 2.6.4 says that $(\hat{\mathbf{h}}_t')_{t\in\mathbb{N}}$ and $(\hat{\mathbf{h}}_t'')_{t\in\mathbb{N}}$ are approximated by $(\mathbf{d}_t)_{t\in\mathbb{N}}$ and $(\tilde{\mathbf{d}}_t)_{t\in\mathbb{N}}$, respectively, with an exponentially fast decaying error. Additional arguments prove that $(\mathbf{d}_t)$ and $(\tilde{\mathbf{d}}_t)$ coincide with $(\mathbf{h}_t')$ and $(\mathbf{h}_t'')$, where we recall that $(\mathbf{h}_t)$ denotes the unique stationary ergodic solution of the SRE $\mathbf{s}_{t+1} = \Phi_t(\mathbf{s}_t)$.

To begin with, we derive a SRE for $(\hat{\mathbf{h}}_t)_{t\in\mathbb{N}}$. We will take the derivative with respect to $\boldsymbol{\theta}$ on both sides of (5.48). For a notationally tractable representation of the arising SREs, we introduce the maps

$$\varphi_t: \quad K \times [0,\infty)^q \to K \times [0,\infty)^q, \quad (\boldsymbol{\theta},\mathbf{u}) \mapsto (\boldsymbol{\theta},(g_{\boldsymbol{\theta}}(\mathbf{X}_t,\mathbf{u}),u_1,\ldots,u_{q-1})),$$

$$p_2: \quad K \times [0,\infty)^q \to [0,\infty)^q, \quad (\boldsymbol{\theta},\mathbf{u}) \mapsto \mathbf{u},$$

and set $\psi_{t,r} = p_2 \circ \varphi_t^{(r)}$ for fixed $r \geq 1$. A comparison with the definition of Φ_t in (5.30) reveals that

$$[\Phi_t^{(r)}(\mathbf{s})](\boldsymbol{\theta}) = \psi_{t,r}(\boldsymbol{\theta}, \mathbf{s}(\boldsymbol{\theta})), \qquad \mathbf{s} \in \mathbb{C}(K, [0,\infty)^q),$$

for every $t \in \mathbb{Z}$ and $r \geq 1$. In this (and only this) section, we work under the convention that the first and second order partial derivatives of a function $\mathbf{f} = (f_1, \dots, f_m)^T : U \subset \mathbb{R}^n \to \mathbb{R}^m$ are written as column vectors:

$$\mathbf{f}' = (\partial_1^1 \mathbf{f}, \partial_1^2 \mathbf{f}, \dots, \partial_1^n \mathbf{f}, \dots, \dots, \partial_m^1 \mathbf{f}, \dots, \partial_m^n \mathbf{f})^T,$$

$$\begin{aligned}\mathbf{f}'' = (&\partial_1^{1,1}\mathbf{f}, \dots, \partial_1^{1,n}\mathbf{f}, \partial_1^{2,1}\mathbf{f}, \dots, \partial_1^{2,n}\mathbf{f}, \dots, \partial_1^{n,1}\mathbf{f}, \dots, \partial_1^{n,n}\mathbf{f},\\ &\partial_2^{1,1}\mathbf{f}, \dots, \partial_2^{n,n}\mathbf{f}, \dots, \partial_m^{1,1}\mathbf{f}, \dots, \partial_m^{n,n}\mathbf{f})^T,\end{aligned}$$

where $\partial_j^k \mathbf{f} := (\partial f_j)/(\partial x_k)$ and $\partial_j^{k_1,k_2}\mathbf{f} := (\partial^2 f_j)/(\partial x_{k_1} \partial x_{k_2})$. In what follows, we suppose that the following regularity conditions hold true:

D.1 The conditions of Proposition 5.2.12 are fulfilled with a compact set $K \subset \Theta \subset \mathbb{R}^d$. Suppose that K coincides with the closure of its (open) interior. The function $(\boldsymbol{\theta}, s) \mapsto g_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{s})$ on $K \times [0,\infty)^q$ is continuously differentiable for every fixed $\mathbf{x} \in \mathbb{R}^p$.

D.2 For all $j \in \{1, \dots, q\}$ and $k \in \{1, \dots, d+q\}$,

$$\mathbb{E}\left[\log^+ \left(\sup_{\boldsymbol{\theta} \in K} |\partial_j^k \psi_{0,1}(\boldsymbol{\theta}, \mathbf{h}_0(\boldsymbol{\theta}))|\right)\right] < \infty. \tag{5.50}$$

Moreover there exist a stationary sequence $(\bar{C}_1(t))$ with $\mathbb{E}[\log^+ \bar{C}_1(0)] < \infty$ and $\kappa \in (0,1]$ such that

$$\sup_{\boldsymbol{\theta} \in K} |\partial_j^k \psi_{t,1}(\boldsymbol{\theta}, \mathbf{u}) - \partial_j^k \psi_{t,1}(\boldsymbol{\theta}, \tilde{\mathbf{u}})| \leq \bar{C}_1(t)\, |\mathbf{u} - \tilde{\mathbf{u}}|^\kappa, \quad \mathbf{u}, \tilde{\mathbf{u}} \in [0,\infty)^q, \tag{5.51}$$

for every $j \in \{1, \dots, q\}$, $k \in \{1, \dots, d+q\}$ and $t \in \mathbb{Z}$.

Equation (5.48) can be understood as

$$\hat{\mathbf{h}}_{t+1}(\boldsymbol{\theta}) = \Phi_t(\hat{\mathbf{h}}_t(\boldsymbol{\theta})) = \psi_{t,1}(\boldsymbol{\theta}, \hat{\mathbf{h}}_t(\boldsymbol{\theta})), \qquad \boldsymbol{\theta} \in K.$$

Taking the derivatives with respect to $\boldsymbol{\theta}$ on both sides yields

$$\partial_j^k \hat{\mathbf{h}}_{t+1}(\boldsymbol{\theta}) = \partial_j^k \psi_{t,1}(\boldsymbol{\theta}, \hat{\mathbf{h}}_t(\boldsymbol{\theta})) + \sum_{i=1}^q \partial_j^{d+i} \psi_{t,1}(\boldsymbol{\theta}, \hat{\mathbf{h}}_t(\boldsymbol{\theta}))\, \partial_i^k \hat{\mathbf{h}}_t(\boldsymbol{\theta}), \qquad t \in \mathbb{N}, \tag{5.52}$$

for indices $j \in \{1, \dots, q\}, k \in \{1, \dots, d\}$, or in abridged form,

$$\hat{\mathbf{h}}'_{t+1} = \hat{\dot{\Phi}}_t(\hat{\mathbf{h}}'_t), \qquad t \in \mathbb{N}.$$

In order to find a stationary approximation to $(\hat{\mathbf{h}}_t')_{t\in\mathbb{N}}$, we replace $\hat{\mathbf{h}}_t$ by $\mathbf{h}_t$ and $\hat{\mathbf{h}}_t'$ by $\mathbf{q}_t$ in (5.52). This leads to the linear SRE

$$\mathbf{q}_{t+1} = \dot{\Phi}_t(\mathbf{q}_t), \qquad t \in \mathbb{Z}, \tag{5.53}$$

on $\mathbb{C}(K, \mathbb{R}^{dq})$, where for $\ell = (j-1)d + k \in \{1, \dots, dq\}$,

$$[\mathbf{q}_{t+1}(\boldsymbol{\theta})]_\ell = \partial_j^k \psi_{t,1}(\boldsymbol{\theta}, \mathbf{h}_t(\boldsymbol{\theta})) + \sum_{i=1}^{q} \partial_j^{d+i} \psi_{t,1}(\boldsymbol{\theta}, \mathbf{h}_t(\boldsymbol{\theta}))\, [\mathbf{q}_t(\boldsymbol{\theta})]_{(i-1)d+k}.$$

We will recognize shortly that the SRE (5.53) is contractive and hence has a unique stationary ergodic solution $(\mathbf{d}_t)$. This indicates that $(\mathbf{d}_t)_{t\in\mathbb{N}}$ provides a stationary approximation to $(\hat{\mathbf{h}}_t')_{t\in\mathbb{N}}$. Note that this rationale is in line with Proposition 5.2.12. In what follows, we show:

(1) The SRE (5.53) has a unique stationary solution $(\mathbf{d}_t)$, which is ergodic. The random element $\mathbf{d}_t$ is $\mathcal{F}_{t-1}$–measurable for every t.

(2) We have that $\|\hat{\mathbf{h}}_t' - \mathbf{d}_t\|_K \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$, i.e., $(\mathbf{d}_t)_{t\in\mathbb{N}}$ is a stationary approximation of $(\hat{\mathbf{h}}_t')_{t\in\mathbb{N}}$.

(3) The random functions $\mathbf{h}_t$ are a.s. continuously differentiable on K, and for each $t \in \mathbb{Z}$,
$$\mathbf{d}_t \equiv \mathbf{h}_t'.$$

Since we establish relation (1) via Theorem 2.6.1, we need to show that $\dot{\Phi}_t^{(r)}$ is a contraction on average for r large enough. As is obvious from elementary calculus,

$$\begin{aligned}[\dot{\Phi}_t^{(r)}(\mathbf{q})(\boldsymbol{\theta})]_\ell &\\ = \partial_j^k \psi_{t,r}(\boldsymbol{\theta}, \mathbf{h}_{t-r+1}(\boldsymbol{\theta})) &+ \sum_{i=1}^{q} \partial_j^{d+i} \psi_{t,r}(\boldsymbol{\theta}, \mathbf{h}_{t-r+1}(\boldsymbol{\theta}))\, [\mathbf{q}(\boldsymbol{\theta})]_{(i-1)d+k}.\end{aligned} \tag{5.54}$$

From $\psi_{t,r}(\boldsymbol{\theta}, \mathbf{u}) = p_2 \circ \varphi_t^{(r)}(\boldsymbol{\theta}, \mathbf{u}) = \Phi_t^{(r)}(\mathbf{u})$ we deduce that

$$|\psi_{t,r}(\boldsymbol{\theta}, \mathbf{u}) - \psi_{t,r}(\boldsymbol{\theta}, \tilde{\mathbf{u}})| \le \Lambda(\Phi_t^{(r)})\, |\mathbf{u} - \tilde{\mathbf{u}}|, \qquad \mathbf{u}, \tilde{\mathbf{u}} \in [0,\infty)^q,$$

and therefore

$$\sup_{\boldsymbol{\theta}\in K} |\partial_j^{d+i} \psi_{t,r}(\boldsymbol{\theta}, \mathbf{h}_{t-r+1}(\boldsymbol{\theta}))| \le \Lambda(\Phi_t^{(r)}), \qquad t \in \mathbb{Z}, \tag{5.55}$$

for all $i \in \{1, \dots, q\}$. Using the representation (5.54) and applying inequality (5.55), we obtain

$$\begin{aligned}\Big| [\dot{\Phi}_t^{(r)}(\mathbf{q})(\boldsymbol{\theta})]_\ell - [\dot{\Phi}_t^{(r)}(\tilde{\mathbf{q}})(\boldsymbol{\theta})]_\ell \Big| &\\ &\le \Lambda(\Phi_t^{(r)}) \sum_{i=1}^{q} \big| [\mathbf{q}(\boldsymbol{\theta})]_{(i-1)d+k} - [\tilde{\mathbf{q}}(\boldsymbol{\theta})]_{(i-1)d+k} \big| \\ &\le \text{const} \times \Lambda(\Phi_t^{(r)}) \, \|\mathbf{q} - \tilde{\mathbf{q}}\|_K, \qquad \mathbf{q}, \tilde{\mathbf{q}} \in \mathbb{C}(K, \mathbb{R}^{dq}),\end{aligned}$$

for all $\ell \in \{1, \dots, dq\}$ and $\boldsymbol{\theta} \in K$, whence

$$\Lambda(\dot{\Phi}_t^{(r)}) \le c\, \Lambda(\Phi_t^{(r)})$$

for a certain constant $c > 0$ not depending on r. It follows from the proof of Theorem 2.6.1 that $\mathbb{E}[\log \Lambda(\Phi_0^{(r)})] \to -\infty$ as $r \to \infty$. For this reason we can choose r so large that

$$\mathbb{E}[\log \Lambda(\dot{\Phi}_0^{(r)})] \le \log c + \mathbb{E}[\log \Lambda(\Phi_0^{(r)})] < 0.$$

Thus the SRE $\mathbf{q}_{t+1} = \dot{\Phi}_t(\mathbf{q}_t)$ obeys the condition S.2 of Theorem 2.6.1, and S.1 is true by virtue of (5.50). Consequently the SRE (5.53) admits a unique stationary ergodic solution $(\mathbf{d}_t)$, for which $\mathbf{d}_t$ is $\mathcal{F}_{t-1}$–measurable for every t. As regards the limit relation (2), we need to study the perturbed SRE

$$\mathbf{q}_{t+1} = \hat{\dot{\Phi}}_t(\mathbf{q}_t), \qquad t \in \mathbb{N},$$

which has $(\hat{\mathbf{h}}_t')_{t\in\mathbb{N}}$ as one of its solutions. By the assumption (5.51), the triangle inequality, Proposition 5.2.12 and an application of Proposition 2.5.1, we have that

$$\|\hat{\dot{\Phi}}_t(\mathbf{0}) - \dot{\Phi}_t(\mathbf{0})\|_K \le \text{const} \times \bar{C}_1(t)\, \|\hat{\mathbf{h}}_t - \mathbf{h}_t\|_K^{\kappa} \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty,$$

and

$$\Lambda(\hat{\dot{\Phi}}_t - \dot{\Phi}_t) \le \text{const} \times \bar{C}_1(t)\, \|\hat{\mathbf{h}}_t - \mathbf{h}_t\|_K^{\kappa} \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty.$$

Now an application of Theorem 2.6.4 demonstrates $\|\hat{\mathbf{h}}_t' - \mathbf{d}_t\|_K \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$. It remains to prove relation (3). Since K coincides with the closure of its interior, the continuous differentiability of $\mathbf{h}_t$ on K can be established by showing the existence of a sequence $(\mathbf{f}_n)_{n\in\mathbb{N}}$ of continuously differentiable functions on K such that $\mathbf{f}_n \xrightarrow{\text{a.s.}} \mathbf{h}_t$ in $\mathbb{C}(K, \mathbb{R}^q)$ and $\mathbf{f}_n' \xrightarrow{\text{a.s.}} \mathbf{d}_t$ in $\mathbb{C}(K, \mathbb{R}^{dq})$ as $n \to \infty$; cf. Theorem 5.9.12 in Lang [80]. For every *fixed* $m \ge 0$, the sequence $((\mathbf{h}_{t,m}', \mathbf{d}_t))_{t\in\mathbb{Z}}$ is stationary ergodic by virtue of Proposition 2.1.1 (see (5.33) for the definition of $\mathbf{h}_{t,m}$). Since $\mathbf{h}_{m,m}' = \hat{\mathbf{h}}_m'$, another application of Proposition 2.1.1 implies

$$\mathbf{h}_{t,m}' - \mathbf{d}_t \stackrel{\text{d}}{=} \hat{\mathbf{h}}_m' - \mathbf{d}_m, \qquad m \ge 0.$$

On the other hand, we have already shown that $\|\hat{\mathbf{h}}_m' - \mathbf{d}_m\|_K \xrightarrow{\text{e.a.s.}} 0$. Thus $\|\mathbf{h}_{t,m}' - \mathbf{d}_t\|_K \xrightarrow{\mathbb{P}} 0$ as $m \to \infty$, and therefore there is a subsequence

$(\mathbf{h}'_{t,m_n})_{n\in\mathbb{N}}$ with $\|\mathbf{h}'_{t,m_n} - \mathbf{d}_t\|_K \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. If we set $\mathbf{f}_n = \mathbf{h}_{t,m_n}$, $n \in \mathbb{N}$, then the sequence $(\mathbf{f}_n)_{n\in\mathbb{N}}$ satisfies $\mathbf{f}_n \xrightarrow{\text{a.s.}} \mathbf{h}_t$ in $\mathbb{C}(K, \mathbb{R}^q)$ and $\mathbf{f}'_n \xrightarrow{\text{a.s.}} \mathbf{d}_t$ in $\mathbb{C}(K, \mathbb{R}^{dq})$ as $n \to \infty$. This completes the proof of assertion (3). Summarizing, we have obtained the following proposition.

Proposition 5.5.1. *Assume that conditions* D.1 *and* D.2 *are fulfilled. Then the* SRE $\mathbf{q}_{t+1} = \dot{\mathbf{\Phi}}_t(\mathbf{q}_t)$ *defined by* (5.53) *has a unique stationary solution* $(\mathbf{d}_t)$, *which is ergodic. For every* $t \in \mathbb{Z}$ *the random element* $\mathbf{d}_t$ *is* $\mathcal{F}_{t-1}$*–measurable. For every* $t \in \mathbb{Z}$, *the first derivatives of* $\mathbf{h}_t$ *coincide with* $\mathbf{s}_t$ *on* K a.s. *Moreover,*

$$\|\hat{\mathbf{h}}'_t - \mathbf{d}_t\|_K \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty.$$

This justifies the following definition for the first derivatives of $\mathbf{h}_t$*:* $\mathbf{h}'_t \equiv \mathbf{d}_t$.

Similar results can be derived for $(\hat{\mathbf{h}}''_t)_{t\in\mathbb{N}}$. At the origin of the analysis we have the observation that $(\hat{\mathbf{h}}''_t)_{t\in\mathbb{N}}$ obeys a SRE, which is contractive provided certain regularity assumptions hold. One can more or less follow the lines of proof of Proposition 5.5.1. In addition to D.1 and D.2, we will also assume the following set of conditions.

D.3 The function $(\boldsymbol{\theta}, \mathbf{s}) \mapsto g_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{s})$ on $K \times [0, \infty)^q$ is twice continuously differentiable for every fixed $\mathbf{x} \in \mathbb{R}^p$. For all $j \in \{1, \ldots, q\}$ and $k_1, k_2 \in \{1, \ldots, d + q\}$,

$$\mathbb{E}\left[\log^+\left(\sup_{\boldsymbol{\theta}\in K} |\partial_j^{k_1,k_2}\psi_{0,1}(\boldsymbol{\theta}, \mathbf{h}_0(\boldsymbol{\theta}))|\right)\right] < \infty. \tag{5.56}$$

The sequence of first derivatives $(\mathbf{h}'_t)$ fulfills $\mathbb{E}(\log^+\|\mathbf{h}'_0\|_K) < \infty$. Moreover there exist a stationary sequence $(\bar{C}_2(t))$ with $\mathbb{E}[\log^+\bar{C}_2(0)] < \infty$ and $\tilde{\kappa} \in (0, 1]$ such that

$$\sup_{\boldsymbol{\theta}\in K} |\partial_j^{k_1,k_2}\psi_{t,1}(\boldsymbol{\theta}, \mathbf{u}) - \partial_j^{k_1,k_2}\psi_{t,1}(\boldsymbol{\theta}, \tilde{\mathbf{u}})| \le \bar{C}_2(t)\,|\mathbf{u} - \tilde{\mathbf{u}}|^{\tilde{\kappa}}, \quad \mathbf{u}, \tilde{\mathbf{u}} \in [0, \infty)^q, \tag{5.57}$$

for every $j \in \{1, \ldots, q\}$, $k_1, k_2 \in \{1, \ldots, d + q\}$ and $t \in \mathbb{Z}$.

Proposition 5.5.2. *Assume that conditions* D.1 – D.3 *are fulfilled. Then*

$$\|\hat{\mathbf{h}}''_t - \tilde{\mathbf{d}}_t\|_K \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty,$$

where $(\tilde{\mathbf{d}}_t)$ *is characterized as the unique stationary ergodic solution of the linear* SRE (5.58) *defined below. For every* $t \in \mathbb{Z}$ *the random element* $\tilde{\mathbf{d}}_t$ *is* $\mathcal{F}_{t-1}$*–measurable. The second derivatives of* $\mathbf{h}_t$ *on* K *coincide with* $\tilde{\mathbf{d}}_t$ a.s. *for every* t. *Therefore the following definition for the second derivatives of* $\mathbf{h}_t$ *is justified:* $\mathbf{h}''_t \equiv \tilde{\mathbf{d}}_t$.

Proof. Differentiation of both sides of (5.52) with respect to $\boldsymbol{\theta}$ shows that $\hat{\mathbf{h}}''_{t+1} = \hat{\ddot{\mathbf{\Phi}}}_t(\hat{\mathbf{h}}''_t)$, where $\hat{\ddot{\mathbf{\Phi}}}_t$ is a linear random map. More precisely, for every $j \in \{1, \ldots, q\}$ and $k_1, k_2 \in \{1, \ldots, d\}$,

$$\partial_j^{k_1,k_2}\,\hat{\mathbf{h}}_{t+1}(\boldsymbol{\theta}) =$$

$$\partial_j^{k_1,k_2}\,\psi_{t,1}(\boldsymbol{\theta},\hat{\mathbf{h}}_t(\boldsymbol{\theta})) + \sum_{i=1}^{q}\partial_j^{k_2,d+i}\,\psi_{t,1}(\boldsymbol{\theta},\hat{\mathbf{h}}_t(\boldsymbol{\theta}))\,\partial_i^{k_1}\,\hat{\mathbf{h}}_t(\boldsymbol{\theta})$$

$$+\sum_{i=1}^{q}\left(\partial_j^{k_1,d+i}\psi_{t,1}(\boldsymbol{\theta},\hat{\mathbf{h}}_t(\boldsymbol{\theta})) + \sum_{i'=1}^{q}\partial_j^{d+i',d+i}\psi_{t,1}(\boldsymbol{\theta},\hat{\mathbf{h}}_t(\boldsymbol{\theta}))\partial_{i'}^{k_1}\hat{\mathbf{h}}_t(\boldsymbol{\theta})\right)\partial_i^{k_2}\,\hat{\mathbf{h}}_t(\boldsymbol{\theta})$$

$$+\sum_{i=1}^{q}\partial_j^{d+i}\,\psi_{t,1}(\boldsymbol{\theta},\hat{\mathbf{h}}_t(\boldsymbol{\theta}))\,\partial_i^{k_1,k_2}\hat{\mathbf{h}}_t(\boldsymbol{\theta}), \qquad t\in\mathbb{N}.$$

This suggests to consider the following linear SRE on $\mathbb{C}(K,\mathbb{R}^{d^2q})$:

$$\mathbf{q}_{t+1} = \ddot{\Phi}_t(\mathbf{q}_t), \qquad t\in\mathbb{Z}, \tag{5.58}$$

where for $\ell = (j-1)d^2 + (k_1-1)d + k_2 \in \{1,\dots,d^2q\}$,

$$[\ddot{\Phi}_t(\mathbf{q})(\boldsymbol{\theta})]_\ell =$$

$$\partial_j^{k_1,k_2}\,\psi_{t,1}(\boldsymbol{\theta},\mathbf{h}_t(\boldsymbol{\theta})) + \sum_{i=1}^{q}\partial_j^{k_2,d+i}\,\psi_{t,1}(\boldsymbol{\theta},\mathbf{h}_t(\boldsymbol{\theta}))\,\partial_i^{k_1}\,\mathbf{h}_t(\boldsymbol{\theta})$$

$$+\sum_{i=1}^{q}\left(\partial_j^{k_1,d+i}\psi_{t,1}(\boldsymbol{\theta},\mathbf{h}_t(\boldsymbol{\theta})) + \sum_{i'=1}^{q}\partial_j^{d+i',d+i}\psi_{t,1}(\boldsymbol{\theta},\mathbf{h}_t(\boldsymbol{\theta}))\partial_{i'}^{k_1}\mathbf{h}_t(\boldsymbol{\theta})\right)\partial_i^{k_2}\,\mathbf{h}_t(\boldsymbol{\theta})$$

$$+\sum_{i=1}^{q}\partial_j^{d+i}\,\psi_{t,1}(\boldsymbol{\theta},\mathbf{h}_t(\boldsymbol{\theta}))\,[\mathbf{q}(\boldsymbol{\theta})]_{(i-1)d^2+(k_1-1)d+k_2}, \qquad \mathbf{q}\in\mathbb{C}(K,\mathbb{R}^{d^2q}).$$

Exploiting the conditions (5.56) and $\mathbb{E}(\log^+\|\mathbf{h}_0'\|_K) < \infty$ in D.3, one shows with exactly the same arguments as in the proof of Proposition 5.5.1 that the SRE (5.58) obeys the conditions of Theorem 2.6.1, which implies that it has a unique stationary ergodic solution $(\tilde{\mathbf{d}}_t)$. Furthermore, $\tilde{\mathbf{d}}_t$ is $\mathcal{F}_{t-1}$–measurable for every t. By means of the decomposition

$$\hat{a}\hat{b}\hat{c} - abc = (\hat{a}-a)bc + (\hat{a}-a)(\hat{b}-b)c + (\hat{a}-a)(\hat{b}-b)(\hat{c}-c)$$
$$+(\hat{a}-a)b(\hat{c}-c) + a(\hat{b}-b)c + a(\hat{b}-b)(\hat{c}-c) + ab(\hat{c}-c) \tag{5.59}$$

and application of the bounds (5.57) together with $\|\hat{\mathbf{h}}_t - \mathbf{h}_t\|_K^{\tilde{\kappa}} \xrightarrow{\text{e.a.s.}} 0$ and $\|\hat{\mathbf{h}}_t' - \mathbf{h}_t'\|_K^{\tilde{\kappa}} \xrightarrow{\text{e.a.s.}} 0$, it can be verified that $(\ddot{\Phi}_t)$ and $(\hat{\ddot{\Phi}}_t)_{t\in\mathbb{N}}$ fulfill the conditions of Theorem 2.6.4. Thus $\|\hat{\mathbf{h}}_t'' - \tilde{\mathbf{d}}_t\|_K \xrightarrow{\text{e.a.s.}} 0$. Analogously to the proof of Proposition 5.5.1 one demonstrates $\tilde{\mathbf{d}}_t \equiv \mathbf{h}_t''$. This concludes the proof. □

5.6 Asymptotic Normality of the QMLE

As we have previously indicated, it is convenient to first establish the asymptotic normality of $\tilde{\boldsymbol{\theta}}_n$, defined in (5.37), and secondly to establish the asymptotic equivalence of $\tilde{\boldsymbol{\theta}}_n$ and the QMLE $\hat{\boldsymbol{\theta}}_n$, i.e., $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n) \xrightarrow{\mathbb{P}} 0$. Following the classical approach, we will establish the asymptotic normality of $\tilde{\boldsymbol{\theta}}_n$ by means of a Taylor expansion of $L_n' = (\sum_{t=1}^n \ell_t)'$, where $\ell_t = -2^{-1}(\log h_t + X_t^2/h_t)$. For this reason it is essential to study the limit properties of L_n' and L_n''. Now we formulate the basic assumptions used throughout this section.

N.1 The assumptions C.1 – C.4 of Section 5.3 are fulfilled and the true parameter $\boldsymbol{\theta}_0$ lies in the interior of the compact set K.

N.2 The assumptions D.1 – D.3 of Proposition 5.5.2 are met for a compact ball $\tilde{K}$ which contains $\boldsymbol{\theta}_0$ in its interior.

N.3 The following moment conditions hold:

(i) $\mathbb{E}Z_0^4 < \infty$,

(ii) $\mathbb{E}\left(\frac{|h_0'(\boldsymbol{\theta}_0)|^2}{\sigma_0^4}\right) < \infty$,

(iii) $\mathbb{E}\|\ell_0''\|_{\tilde{K}} < \infty$.

N.4 The components of the vector $\frac{\partial g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}(X_0, \sigma_0^2)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ are linearly independent random variables.

By virtue of Theorem 5.3.1, condition N.1 implies consistency. The requirement N.2 enables the differentiation of L_n on $\tilde{K}$. Since $\tilde{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$, only the differentiability of L_n in a neighborhood of $\boldsymbol{\theta}_0$ is needed in the proof. For the sake of simpler arguments we assume without loss of generality that $\tilde{K} = K$. Assumption N.4 assures that the asymptotic covariance matrix is regular.

Observe that with probability one,

$$\ell_t'(\boldsymbol{\theta}) = -\frac{1}{2}\frac{h_t'(\boldsymbol{\theta})}{h_t(\boldsymbol{\theta})}\left(1 - \frac{X_t^2}{h_t(\boldsymbol{\theta})}\right), \tag{5.60}$$

$$\ell_t''(\boldsymbol{\theta}) = -\frac{1}{2}\frac{1}{h_t(\boldsymbol{\theta})^2}\left((h_t'(\boldsymbol{\theta}))^T h_t'(\boldsymbol{\theta})\left(2\frac{X_t^2}{h_t(\boldsymbol{\theta})} - 1\right) + h_t''(\boldsymbol{\theta})(h_t(\boldsymbol{\theta}) - X_t^2)\right). \tag{5.61}$$

From Propositions 5.2.12, 5.5.1 and 5.5.2 together with Proposition 2.1.1 we infer that (ℓ_t') and (ℓ_t'') are stationary ergodic sequences of random elements with values in $\mathbb{C}(K, \mathbb{R}^d)$ and $\mathbb{C}(K, \mathbb{R}^{d\times d})$, respectively. An inspection of the proof of Theorem 5.3.1 shows that condition N.1 also implies $\tilde{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$. Consequently, for large enough n the following Taylor expansion is valid:

$$L'_n(\tilde{\boldsymbol{\theta}}_n) = L'_n(\boldsymbol{\theta}_0) + L''_n(\boldsymbol{\zeta}_n)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \tag{5.62}$$

where $|\boldsymbol{\zeta}_n - \boldsymbol{\theta}_0| < |\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|$. Since $\tilde{\boldsymbol{\theta}}_n$ is the maximizer of L_n and $\boldsymbol{\theta}_0$ lies in the interior of K, one has $L'_n(\tilde{\boldsymbol{\theta}}_n) = \mathbf{0}$. Therefore (5.62) is equivalent to

$$n^{-1}L''_n(\boldsymbol{\zeta}_n)\,(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -n^{-1}L'_n(\boldsymbol{\theta}_0). \tag{5.63}$$

On account of $\mathbb{E}\|\ell''_0\|_K < \infty$ and the stationarity and ergodicity of (ℓ''_t), we may apply Theorem 2.2.1 to obtain $L''_n/n \xrightarrow{\text{a.s.}} L''$ in $\mathbb{C}(K, \mathbb{R}^{d\times d})$ as $n \to \infty$, where $L''(\boldsymbol{\theta}) = \mathbb{E}[\ell''_0(\boldsymbol{\theta})]$, $\boldsymbol{\theta} \in K$. This uniform convergence result together with $\boldsymbol{\zeta}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$ implies

$$L''_n(\boldsymbol{\zeta}_n)/n \xrightarrow{\text{a.s.}} \mathbb{E}[\ell''_0(\boldsymbol{\theta}_0)] = \mathbf{F}_0, \qquad n \to \infty.$$

By Propositions 5.2.12, 5.5.1 and 5.5.2 h_0, h'_0 and h''_0 are $\mathcal{F}_{-1}$–measurable. Exploiting $h_0(\boldsymbol{\theta}_0) = \sigma_0^2$ a.s., $X_0 = \sigma_0 Z_0$ and the independence of Z_0 and $\mathcal{F}_{-1}$, one may conclude that

$$\mathbf{F}_0 = -2^{-1}\mathbb{E}\left[(h'_0(\boldsymbol{\theta}_0))^T h'_0(\boldsymbol{\theta}_0)/\sigma_0^4\right]. \tag{5.64}$$

It is shown in Lemma 5.6.3 below that $\mathbf{F}_0$ is invertible. Consequently, the matrix $L''_n(\boldsymbol{\zeta}_n)/n$ has inverse $\mathbf{F}_0^{-1}(1 + o_{\mathbb{P}}(1))$, $n \to \infty$, and (5.63) implies

$$\sqrt{n}\,(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\mathbf{F}_0^{-1}(1 + o_{\mathbb{P}}(1))L'_n(\boldsymbol{\theta}_0)/\sqrt{n}, \qquad n \to \infty. \tag{5.65}$$

Therefore the limit of $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is determined by that of $-\mathbf{F}_0^{-1}L'_n(\boldsymbol{\theta}_0)/\sqrt{n}$. Since $h_t(\boldsymbol{\theta}_0) = \sigma_t^2$ a.s. and $X_t = \sigma_t Z_t$,

$$L'_n(\boldsymbol{\theta}_0) = \sum_{t=1}^{n} \ell'_t(\boldsymbol{\theta}_0) = \frac{1}{2}\sum_{t=1}^{n} \frac{h'_t(\boldsymbol{\theta}_0)}{\sigma_t^2}(Z_t^2 - 1).$$

Since the random element h'_t/σ_t^2 is $\mathcal{F}_{t-1}$–measurable and since $\mathcal{F}_{t-1}$ is independent of Z_t and $\mathbb{E}Z_t^2 = 1$, the sequence $(\ell'_t(\boldsymbol{\theta}_0))_{t\in\mathbb{N}}$ is a stationary ergodic zero–mean martingale difference sequence with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$. By virtue of the moment condition N.3 the sequence $(\ell'_t(\boldsymbol{\theta}_0))_{t\in\mathbb{N}}$ has furthermore finite variance. Consequently we can apply the central limit theorem for finite variance stationary ergodic martingale difference sequences, cf. Theorem 18.3 in Billingsley [13], which says that

$$n^{-1/2}L'_n(\boldsymbol{\theta}_0) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, \mathbf{G}_0), \qquad n \to \infty,$$

where

$$\mathbf{G}_0 = \mathbb{E}[(\ell'_0(\boldsymbol{\theta}_0))^T \ell'_0(\boldsymbol{\theta}_0)] = 4^{-1}\mathbb{E}(Z_0^4 - 1)\,\mathbb{E}[(h'_0(\boldsymbol{\theta}_0))^T h'_0(\boldsymbol{\theta}_0)/\sigma_0^4].$$

Together with (5.64) and (5.65) we conclude

$$\sqrt{n}\,(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0}, \mathbf{V}_0), \qquad n \to \infty,$$

where

$$\mathbf{V}_0 = \mathbf{F}_0^{-1}\mathbf{G}_0\mathbf{F}_0^{-1} = 4^{-1}\mathbb{E}(Z_0^4 - 1)\,(\mathbb{E}[(h_0'(\boldsymbol{\theta}_0))^T h_0'(\boldsymbol{\theta}_0)/\sigma_0^4])^{-1}. \tag{5.66}$$

It is shown in Lemma 5.6.5 below that $\sqrt{n}\,|\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n| \xrightarrow{\text{a.s.}} 0$ so that an application of Slutsky's lemma finalizes the proof of the following theorem.

Theorem 5.6.1. *Under the conditions* N.1 – N.4*, the* QMLE $\hat{\boldsymbol{\theta}}_n$ *is strongly consistent and asymptotically normal, i.e.,*

$$\sqrt{n}\,(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0}, \mathbf{V}_0), \qquad n \to \infty,$$

where the asymptotic covariance matrix $\mathbf{V}_0$ *is given by* (5.66).

Remark 5.6.2. In general it seems impossible to find a tractable expression for the asymptotic covariance matrix $\mathbf{V}_0$ due to the fact that the joint distribution of $(\sigma_0^2, h_0'(\boldsymbol{\theta}_0))$ is not known, not even for GARCH(1, 1). It is however possible to consistently estimate $\mathbf{V}_0$ from the data. Defining the residual by $\hat{Z}_t^{(n)} = X_t/(\hat{h}_t(\hat{\boldsymbol{\theta}}_n))^{1/2}$, the matrix sequence

$$\hat{\mathbf{V}}_0^{(n)} = \left(\frac{1}{4n}\sum_{t=1}^{n}((\hat{Z}_t^{(n)})^4 - 1)\right)\left(\frac{1}{n}\sum_{t=1}^{n}\frac{(\hat{h}_t'(\hat{\boldsymbol{\theta}}_n))^T\hat{h}_t'(\hat{\boldsymbol{\theta}}_n)}{\hat{h}_t(\hat{\boldsymbol{\theta}}_n)^2}\right)^{-1}$$

is a strongly consistent estimator for the matrix $\mathbf{V}_0$. We sketch how this can be demonstrated. Define

$$\mathbf{M}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\frac{(h_t'(\boldsymbol{\theta}))^T h_t'(\boldsymbol{\theta})}{(h_t(\boldsymbol{\theta}))^2} \quad \text{and} \quad \hat{\mathbf{M}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\frac{(\hat{h}_t'(\boldsymbol{\theta}))^T\hat{h}_t'(\boldsymbol{\theta})}{(\hat{h}_t(\boldsymbol{\theta}))^2}$$

and suppose $\mathbb{E}\|\,h_0'/h_0\|_K^2 < \infty$ in addition to N.1 – N.4. Then by an application of Theorem 2.2.1, $\mathbf{M}_n \xrightarrow{\text{a.s.}} \mathbf{M}$ in $\mathbb{C}(K, \mathbb{R}^{d\times d})$, where $\mathbf{M}(\boldsymbol{\theta}) = \mathbb{E}[(h_0'(\boldsymbol{\theta}))^T h_0'(\boldsymbol{\theta})/(h_0(\boldsymbol{\theta}))^2]$. Using the same method as in Lemma 5.6.4, one derives a bound for $\|\hat{\mathbf{M}}_n - \mathbf{M}_n\|_K$ and shows $\|\hat{\mathbf{M}}_n - \mathbf{M}_n\|_K \xrightarrow{\text{a.s.}} 0$, which implies $\hat{\mathbf{M}}_n \xrightarrow{\text{a.s.}} \mathbf{M}$ in $\mathbb{C}(K, \mathbb{R}^{d\times d})$ as $n \to \infty$. Therefore $(\hat{\mathbf{M}}_n(\hat{\boldsymbol{\theta}}_n))^{-1} \xrightarrow{\text{a.s.}} (\mathbf{M}(\boldsymbol{\theta}_0))^{-1}$. Likewise, $n^{-1}\sum_{t=1}^{n}((\hat{Z}_t^{(n)})^4 - 1)$ can be shown to converge to $\mathbb{E}(Z_0^4 - 1)$ a.s. Altogether, $\hat{\mathbf{V}}_0^{(n)} \xrightarrow{\text{a.s.}} \mathbf{V}_0$. We also mention that the practical implementation for the computation of the matrix $\hat{\mathbf{V}}_0^{(n)}$ becomes particularly simple if one makes use of the recursion for $\hat{\mathbf{h}}_t'$, see equation (5.52). □

For the proof of the remaining assertions, we first show that the asymptotic covariance matrix is regular.

Lemma 5.6.3. *The assumptions* N.1 – N.4 *imply that* $\mathbf{F}_0 = \mathbb{E}[\ell_0''(\boldsymbol{\theta}_0)]$ *is negative definite.*

Proof. $\mathbf{F}_0$ being negative definite is equivalent to $\mathbf{J}_0 = \mathbb{E}[(h_0'(\boldsymbol{\theta}_0))^T h_0'(\boldsymbol{\theta}_0)/\sigma_0^4]$ being positive definite. It is evident that $\mathbf{J}_0$ is positive semi–definite. Assume $\mathbf{x}_0^T \mathbf{J}_0 \mathbf{x}_0 = 0$, some $\mathbf{x}_0 \in \mathbb{R}^d$. This is equivalent to

$$\mathbb{E}\left|\frac{h_0'(\boldsymbol{\theta}_0)\mathbf{x}_0}{\sigma_0^2}\right|^2 = 0,$$

which implies $h_0'(\boldsymbol{\theta}_0)\mathbf{x}_0 = 0$ a.s. Note also that $\mathbf{h}_t'(\boldsymbol{\theta}_0)\mathbf{x}_0 = 0$ for every t due to the stationarity of $(\mathbf{h}_t')$. Differentiation of $h_1(\boldsymbol{\theta}) = g_{\boldsymbol{\theta}}(\mathbf{X}_0, \mathbf{h}_0(\boldsymbol{\theta}))$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ yields

$$h_1'(\boldsymbol{\theta}_0) = \frac{\partial g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}(\mathbf{X}_0, \boldsymbol{\sigma}_0^2)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \frac{\partial g_{\boldsymbol{\theta}}}{\partial \mathbf{s}}(\mathbf{X}_0, \boldsymbol{\sigma}_0^2)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \mathbf{h}_0'(\boldsymbol{\theta}_0).$$

Multiplying this equation from the right with $\mathbf{x}_0$ and accounting for $h_1'(\boldsymbol{\theta}_0)\mathbf{x}_0 = 0$ and $\mathbf{h}_0'(\boldsymbol{\theta}_0)\mathbf{x}_0 = \mathbf{0}$ results in

$$\frac{\partial g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}(\mathbf{X}_0, \boldsymbol{\sigma}_0^2)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \mathbf{x}_0 = 0 \quad \text{a.s.}$$

Condition N.4 implies $\mathbf{x}_0 = \mathbf{0}$. This concludes the proof. □

The remaining steps are devoted to the proof of Lemma 5.6.5.

Lemma 5.6.4. *The assumptions* N.1 – N.2 *imply*

$$n^{-1/2}\,\|\hat{L}_n' - L_n'\|_K \xrightarrow{\text{a.s.}} 0, \qquad n \to \infty.$$

Proof. Notice that C.3 implies $\hat{h}_t(\boldsymbol{\theta}), h_t(\boldsymbol{\theta}) \geq \underline{g} > 0$ for all $\boldsymbol{\theta} \in K$. This and the mean value theorem applied to the function $f(a,b) = ab^{-1}(1 - X_t^2/b)$, $a \in \mathbb{R}$, $b \geq \underline{g}$, yield

$$\begin{aligned}\|\hat{\ell}_t' - \ell_t'\|_K &= \left\|\frac{\hat{h}_t'}{\hat{h}_t}\left(1 - \frac{X_t^2}{\hat{h}_t}\right) - \frac{h_t'}{h_t}\left(1 - \frac{X_t^2}{h_t}\right)\right\|_K \\ &\leq c(1 + X_t^2)\left\{\|\hat{h}_t - h_t\|_K + \|\hat{h}_t' - h_t'\|_K + \|\hat{h}_t' - h_t'\|_K^2\,\|h_t'\|_K\right\},\end{aligned} \tag{5.67}$$

for some $c > 0$. Recall that we assume $\mathbb{E}(\log^+ \sigma_0^2) < \infty$, $\mathbb{E}(\log^+ \|h_0'\|_K) < \infty$, and observe that Propositions 5.2.12 and 5.5.1 imply $\|\hat{h}_t - h_t\|_K \xrightarrow{\text{e.a.s.}} 0$ and $\|\hat{h}_t' - h_t'\|_K \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$. Now (5.67) together with an application of Proposition 2.5.1 and Lemma 2.5.3 show $\|\hat{L}_n' - L_n'\|_K \leq \sum_{t=1}^{\infty} \|\hat{\ell}_t' - \ell_t'\|_K < \infty$ a.s. This completes the proof. □

Lemma 5.6.5. *The assumptions* N.1 – N.4 *imply*

$$\sqrt{n}\,|\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n| \xrightarrow{\text{a.s.}} 0, \qquad n \to \infty.$$

Proof. From the mean value theorem,

$$L_n'(\tilde{\boldsymbol{\theta}}_n) - L_n'(\hat{\boldsymbol{\theta}}_n) = L_n''(\tilde{\boldsymbol{\zeta}}_n)(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n), \tag{5.68}$$

where $\tilde{\boldsymbol{\zeta}}_n$ lies on the line segment connecting $\hat{\boldsymbol{\theta}}_n$ and $\tilde{\boldsymbol{\theta}}_n$. This line segment is completely contained in the interior of K provided n is large enough. Since $L_n'(\tilde{\boldsymbol{\theta}}_n) = \hat{L}_n'(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}$, equation (5.68) is equivalent to

$$n^{-1/2}(\hat{L}_n'(\hat{\boldsymbol{\theta}}_n) - L_n'(\hat{\boldsymbol{\theta}}_n)) = n^{-1} L_n''(\tilde{\boldsymbol{\zeta}}_n)\, n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n). \tag{5.69}$$

By virtue of Lemma 5.6.4, both sides of (5.69) tend to $\mathbf{0}$ a.s. as $n \to \infty$. Because of N.3 (iii) we can apply Theorem 2.2.1 to L_n''/n and together with $\tilde{\boldsymbol{\zeta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$ conclude $L_n''(\tilde{\boldsymbol{\zeta}}_n)/n \xrightarrow{\text{a.s.}} \mathbf{F}_0 = \mathbb{E}[\ell_0''(\boldsymbol{\theta}_0)]$. Since the matrix $\mathbf{F}_0$ is invertible, as shown by Lemma 5.6.3, we can deduce $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n) \xrightarrow{\text{a.s.}} \mathbf{0}$, which completes the proof. □

5.7 Examples: Asymptotic Normality

In this section we establish the asymptotic normality of the QMLE in AGARCH(p, q) and EGARCH models. We start with the simpler case of AGARCH(p, q). For EGARCH we can only treat the special case of models with $\beta = 0$.

5.7.1 AGARCH(p,q)

Take a compact set $K \subset (0, \infty) \times [0, \infty)^p \times B \times [-1, 1]$, where $B = \{(\beta_1, \dots, \beta_q)^T \in [0, 1)^q \mid \sum_{j=1}^q \beta_j < 1\}$. Assume that the true parameter vector $\boldsymbol{\theta}_0 = (\alpha_0^\circ, \dots, \alpha_p^\circ, \beta_1^\circ, \dots, \beta_q^\circ, \gamma^\circ)^T$ is contained in the interior of K and suppose the conditions of Theorem 5.4.6 hold true. This entails N.1. For the verification of N.2 – N.3 we can assume without loss of generality that K is a compact ball about $\boldsymbol{\theta}_0$ with $\alpha_i, \beta_j > 0$ for all i, j and $|\gamma| < 1$ on K. Analogously to Berkes et al. [8] we assume $\mathbb{E}Z_0^4 < \infty$ and suppose there is $\mu > 0$ such that

$$\mathbb{P}(|Z_0| \le z) = o(z^\mu), \qquad z \downarrow 0. \tag{5.70}$$

As regards N.2, the verification of conditions D.1 – D.3 of Section 5.5 is rather straightforward. The only steps which require some care are the moment conditions. We have shown in Example 5.2.5 that there is an $\eta > 0$ with $\mathbb{E}|\sigma_0|^{2\eta} < \infty$, and hence also $\mathbb{E}|X_0|^{2\eta} < \infty$ (provided $\eta \le 1$). The Minkowski inequality applied to (5.43) yields $\mathbb{E}\|h_0\|_K^\eta < \infty$. Altogether $\mathbb{E}(\log^+ \sigma_0^2) < \infty$ and $\mathbb{E}(\log^+ \|h_0\|_K) < \infty$. As to the moments of h_t', observe that differentiation (with respect to $\boldsymbol{\theta}$) and the sum in (5.43) can be interchanged. Indeed, the equality

$$\pi_\ell'(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{1}{\ell!} \frac{\partial^\ell}{\partial z^\ell} \left(\frac{\alpha_{\boldsymbol{\theta}}(z)}{\beta_{\boldsymbol{\theta}}(z)} \right) \right) \Big|_{z=0} = \frac{1}{\ell!} \frac{\partial^\ell}{\partial z^\ell} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{\alpha_{\boldsymbol{\theta}}(z)}{\beta_{\boldsymbol{\theta}}(z)} \right) \right) \Big|_{z=0}$$

shows how to compute $\pi'_\ell(\boldsymbol{\theta})$ from $\alpha_{\boldsymbol{\theta}}(z)/\beta_{\boldsymbol{\theta}}(z)$. An application of the Cauchy inequalities to $\partial/(\partial\boldsymbol{\theta})(\alpha_{\boldsymbol{\theta}}(z)/\beta_{\boldsymbol{\theta}}(z))$, where one exploits the fact that $\beta_{\boldsymbol{\theta}}(z) \neq 0$ for $|z| \leq \bar{\beta}^{-1/q}$ (see the proof of Lemma 5.4.3), demonstrates that there are $0 < \tilde{\lambda} < 1$ and $\tilde{c} > 0$ with $\|\pi'_\ell\|_K \leq \tilde{c}\tilde{\lambda}^\ell$. This together with Proposition 2.5.1 shows that the sequence of first derivatives of $f_m(\boldsymbol{\theta}) := \pi_0(\boldsymbol{\theta}) + \sum_{\ell=1}^m \pi_\ell(\boldsymbol{\theta})(|X_{t-\ell}| - \gamma X_{t-\ell})^2$ converges a.s. on $\mathbb{C}(K, \mathbb{R}^{p+q+2})$ as $m \to \infty$. Since with probability one $f_m \xrightarrow{\text{a.s.}} h_t$ in $\mathbb{C}(K)$ and (f'_m) converges uniformly on K,

$$h'_t(\boldsymbol{\theta}) = \pi'_0(\boldsymbol{\theta}) + \sum_{\ell=1}^{\infty} \pi'_\ell(\boldsymbol{\theta})\,(|X_{t-\ell}| - \gamma X_{t-\ell})^2$$

$$-2\Big(\sum_{\ell=1}^{\infty} \pi_\ell(\boldsymbol{\theta}) X_{t-\ell}(|X_{t-\ell}| - \gamma X_{t-\ell})\Big)\mathbf{e}_{p+q+2} \quad \text{a.s.,} \qquad (5.71)$$

where $\mathbf{e}_{p+q+2} = (0,\dots,0,1)^T \in \mathbb{R}^{p+q+2}$. This establishes that differentiation and summation in (5.43) can be interchanged. The representation (5.71) together with the identical arguments for $\mathbb{E}\|h_0\|_K^\eta < \infty$ yields $\mathbb{E}\|h'_0\|_K^\eta < \infty$. Consequently $\mathbb{E}(\log^+ \|h'_0\|_K) < \infty$. Altogether we have established N.2.

It is more complicated to prove the moment conditions of N.3. At the origin lies the observation that the random variables $\|h'_0/h_0\|_K$ and $\|h''_0/h_0\|_K$ have finite moments of *any* order and that

$$\mathbb{E}\|X_0^2/h_0\|_K^\kappa < \infty \qquad (5.72)$$

for any $\kappa < 2$; see Lemmas 5.7.4 and 5.7.5 below. The inequality (2.11) together with the triangle inequality applied to (5.61) imply

$$\|\ell''_0\|_K \leq \frac{1}{2}\left(\left\|\frac{h'_0}{h_0}\right\|_K^2 \left(2\left\|\frac{X_0^2}{h_0}\right\|_K + 1\right) + \left\|\frac{h''_0}{h_0}\right\|_K \left(1 + \left\|\frac{X_0^2}{h_0}\right\|_K\right)\right) \qquad (5.73)$$

By an application of the Hölder inequality together with Lemmas 5.7.4 and 5.7.5, $\mathbb{E}\|\ell''_0\|_K < \infty$ follows. Thus conditions N.1 – N.3 of Section 5.6 are fulfilled.

N.4 will be verified in Lemma 5.7.3 below. Now an application of Theorem 5.6.1 yields the following result.

Theorem 5.7.1. *Let (X_t) be a stationary* AGARCH(p,q) *process with parameter vector $\boldsymbol{\theta}_0 = (\alpha_0^\circ,\dots,\alpha_p^\circ,\beta_1^\circ,\dots,\beta_q^\circ,\gamma^\circ)^T$ such that the polynomials*

$$\alpha^\circ(z) = \sum_{i=1}^{p} \alpha_i^\circ z^i \quad \textit{and} \quad \beta^\circ(z) = 1 - \sum_{j=1}^{q} \beta_j^\circ z^j$$

do not have any common roots. Let $K \subset (0,\infty) \times [0,\infty)^p \times B \times [-1,1]$ be compact and contain $\boldsymbol{\theta}_0$ in its interior. Assume that the distribution of Z_0

is not concentrated in two points and that Z_0 *fulfills* $\mathbb{E}Z_0^4 < \infty$ *and* (5.70). *Then the* QMLE (5.35) *is strongly consistent and asymptotically normal with asymptotic covariance matrix* $\mathbf{V}_0$ *given by* (5.66).

Remarks 5.7.2.

1. Note that the requirement that $\boldsymbol{\theta}_0$ lies in the interior of K entails $\alpha_i^\circ > 0$ for some $i > 0$ and $(\alpha_p^\circ, \beta_q^\circ) \neq (0,0)$, i.e., all conditions of Theorem 5.4.6 are met under the assumptions of Theorem 5.7.1.

2. Mutatis mutandis Theorem 5.7.1 is true also for GARCH(p,q); cf. Theorem 4.2.1. The necessary notational modifications and changes in proofs are obvious.

Lemma 5.7.3. *Under the conditions imposed by Theorem 5.7.1, the condition* N.4 *is fulfilled.*

Proof. By straightforward computation,

$$\begin{aligned}
&\frac{\partial g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}(\mathbf{X}_0, \boldsymbol{\sigma}_0^2)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= (1, (|X_0| - \gamma^\circ X_0)^2, \ldots, (|X_{-p+1}^2| - \gamma^\circ X_{-p+1})^2, \sigma_0^2, \ldots, \sigma_{-q+1}^2, \\
&\quad -2\alpha_1^\circ X_0(|X_0| - \gamma^\circ X_0) - \cdots - 2\alpha_p^\circ X_{-p+1}(|X_{-p+1}| - \gamma^\circ X_{-p+1}))^T.
\end{aligned}$$

Assume for a $\boldsymbol{\xi} = (\lambda_0, \ldots, \lambda_p, \mu_1, \ldots, \mu_q, \nu)^T \in \mathbb{R}^{p+q+2}$ that

$$\frac{\partial g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}(\mathbf{X}_0, \boldsymbol{\sigma}_0^2)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \boldsymbol{\xi} = 0. \tag{5.74}$$

Writing out the latter equation results in

$$\lambda_0 + \sigma_0^2 \left(\lambda_1(|Z_0| - \gamma^\circ Z_0)^2 + \mu_1 - 2\nu\alpha_1^\circ Z_0(|Z_0| - \gamma^\circ Z_0)\right) + Y_{-1} = 0 \qquad \text{a.s.},$$

where Y_{-1} is a certain $\mathcal{F}_{-1}$–measurable random variable. By Lemma 5.4.2 the multiplier of σ_0^2 in the latter identity must be degenerate, i.e.,

$$(\lambda_1(1 + (\gamma^\circ)^2) + 2\nu\alpha_1^\circ\gamma^\circ)Z_0^2 - (2\gamma^\circ\lambda_1 + 2\nu\alpha_1^\circ)Z_0|Z_0| + \mu_1 = c \quad \text{a.s.} \tag{5.75}$$

for a certain $c \in \mathbb{R}$. Since the distribution of Z_0 is not concentrated in two points, $\{1, Z_0|Z_0|, Z_0^2\}$ are linearly independent. Combining this information with (5.75) and taking into account that $\alpha_1^\circ > 0$ and $|\gamma^\circ| < 1$, we conclude $\lambda_1 = \nu = \mu_1 - c = 0$. Thus we must show the linear independence of $1, (|X_0| - \gamma^\circ X_0)^2, \ldots, (|X_{-p+1}^2| - \gamma^\circ X_{-p+1})^2, \sigma_0^2, \ldots, \sigma_{-q+1}^2$. The following arguments are similar to the ones used by Berkes et al. [8] for the proof of their Lemma 5.7. Equation (5.74) with $\nu = 0$ and the stationarity of $((X_t, \sigma_t))$ imply

$$\lambda_0 + \sum_{i=1}^{p} \lambda_i(|X_{t-i}| - \gamma^\circ X_{t-i})^2 + \sum_{j=1}^{q} \mu_j \sigma_{t-j}^2 = 0 \qquad \text{a.s.},$$

which equivalently written in backshift operator notation becomes

$$\lambda_0 + \lambda(B)(|X_t| - \gamma^\circ X_t)^2 + \mu(B)\sigma_t^2 = 0 \qquad \text{a.s.,} \tag{5.76}$$

with

$$\lambda(z) = \sum_{i=1}^{p} \lambda_i z^i \quad \text{and} \quad \mu(z) = \sum_{j=1}^{q} \mu_j z^j.$$

Recall Lemma 5.4.3, where we derived the a.s. representation

$$\sigma_t^2 = \frac{\alpha_0}{\beta^\circ(1)} + \alpha^\circ(B)(\beta^\circ(B))^{-1}(|X_t| - \gamma^\circ X_t)^2 \tag{5.77}$$

with $\alpha^\circ(z) = \sum_{i=1}^{p} \alpha_i^\circ z^i$ and $\beta^\circ(z) = 1 - \sum_{j=1}^{q} \beta_j^\circ z^j$. Plugging (5.77) into (5.76), we obtain that

$$\lambda_0 + \frac{\mu(1)\alpha_0^\circ}{\beta^\circ(1)} + \left(\lambda(B) + \mu(B)\alpha^\circ(B)(\beta^\circ(B))^{-1}\right)(|X_t| - \gamma^\circ X_t)^2 = 0 \qquad \text{a.s.,}$$

and with the identical arguments as given in Lemma 5.4.5, we conclude

$$\lambda(z) + \mu(z)\frac{\alpha^\circ(z)}{\beta^\circ(z)} \equiv 0 \quad \text{and} \quad \lambda_0 + \frac{\mu(1)\alpha_0^\circ}{\beta^\circ(1)} \equiv 0.$$

If $\mu \neq 0$, the rational function $\mu(z)\alpha^\circ(z)/\beta^\circ(z)$ has at least $p+1$ zeros (counted with their multiplicities) since $\mu(0) = 0$ and $\beta^\circ(0) > 0$, whereas $\lambda(z)$ has at most p roots, which is a contradiction. Therefore $\mu(z) = \lambda(z) = 0$ and $\boldsymbol{\xi} = 0$, which completes the proof. □

In what follows, we formulate the AGARCH counterparts of Lemmas 5.1 and 5.2 in Berkes et al. [8].

Lemma 5.7.4. *Let (X_t) be a stationary* AGARCH(p, q) *process with parameter vector $\boldsymbol{\theta}_0 = (\alpha_0^\circ, \ldots, \alpha_p^\circ, \beta_1^\circ, \ldots, \beta_q^\circ, \gamma^\circ)^T$. Let $K \subset (0, \infty) \times [0, \infty)^p \times B \times [-1, 1]$ be a compact ball about $\boldsymbol{\theta}_0$ such that $\alpha_i, \beta_j > 0$ for all i, j and $|\gamma| < 1$ on K. Assume that the iid innovations (Z_t) are such that $\mathbb{E}|Z_0|^{2\tilde{\kappa}} < \infty$ and* (5.70) *holds. Then for any $0 < \kappa < \tilde{\kappa}$,*

$$\mathbb{E}\left\|\frac{X_0^2}{h_0}\right\|_K^\kappa < \infty. \tag{5.78}$$

Proof. In the proof of Lemma 5.1 in Berkes et al. [8] one has to replace (the innovations) (ϵ_t) by $(|Z_t| - \gamma Z_t)$ and (the observations) (y_t) by $(|X_t| - \gamma X_t)$. Also take into account that there is $c_1 > 0$ such that for all $\boldsymbol{\theta} \in K$,

$$c_1 Z_t^2 \le (|Z_t| - \gamma Z_t)^2 \le 2Z_t^2 \quad \text{and} \quad c_1 X_t^2 \le (|X_t| - \gamma X_t)^2 \le 2X_t^2, \tag{5.79}$$

since $|\gamma| < 1$ on K. Exploiting this observation in the steps of proof by Berkes et al. [8] establishes the assertion (5.78). We omit the details. □

Lemma 5.7.5. *Let (X_t) be a stationary* AGARCH(p,q) *process with parameter vector $\boldsymbol{\theta}_0 = (\alpha_0^\circ, \ldots, \alpha_p^\circ, \beta_1^\circ, \ldots, \beta_q^\circ, \gamma^\circ)^T$. Let $K \subset (0,\infty) \times [0,\infty)^p \times B \times [-1,1]$ be a compact ball about $\boldsymbol{\theta}_0$ such that $\alpha_i, \beta_j > 0$ for all i,j and $|\gamma| < 1$ on K. Then for any $\kappa > 0$,*

$$\mathbb{E}\left\|\frac{h_0'}{h_0}\right\|_K^\kappa < \infty \quad \text{and} \quad \mathbb{E}\left\|\frac{h_0''}{h_0}\right\|_K^\kappa < \infty \tag{5.80}$$

Proof. The arguments of Lemma 3.2 in Berkes et al. [8] show that there exist $c_2, c_3 > 0$ with

$$|\pi_\ell'(\boldsymbol{\theta})| \le c_2 \ell \, \pi_\ell(\boldsymbol{\theta}) \quad \text{and} \quad \|\pi_\ell''(\boldsymbol{\theta})\| \le c_3 \ell^2 \, \pi_\ell(\boldsymbol{\theta}), \qquad \ell \ge 1 \tag{5.81}$$

for all $\boldsymbol{\theta} \in K$. Differentiating both sides of (5.71) with respect to $\boldsymbol{\theta}$ yields

$$\begin{aligned} h_t''(\boldsymbol{\theta}) = \pi_0''(\boldsymbol{\theta}) &+ \sum_{\ell=1}^\infty \pi_\ell''(\boldsymbol{\theta}) \, (|X_{t-\ell}| - \gamma X_{t-\ell})^2 \\ &-2\Big(\sum_{\ell=1}^\infty (\pi_\ell'(\boldsymbol{\theta}))^T X_{t-\ell}(|X_{t-\ell}| - \gamma X_{t-\ell})\Big) \mathbf{e}_{p+q+2}^T \\ &-2\mathbf{e}_{p+q+2}\Big(\sum_{\ell=1}^\infty \pi_\ell'(\boldsymbol{\theta}) X_{t-\ell}(|X_{t-\ell}| - \gamma X_{t-\ell})\Big) \\ &+2\Big(\sum_{\ell=1}^\infty \pi_\ell(\boldsymbol{\theta}) X_{t-\ell}^2\Big) \mathbf{e}_{p+q+2}\mathbf{e}_{p+q+2}^T. \end{aligned}$$

This together with (5.81) and applications of (5.79) proves that there is a constant $C > 0$ with

$$|h_t'(\boldsymbol{\theta})|, \ \|h_t''(\boldsymbol{\theta})\| \le C\Big(1 + \sum_{\ell=1}^\infty \ell^2 \pi_\ell(\boldsymbol{\theta}) X_{t-\ell}^2\Big).$$

With this and (5.79) one can follow the lines of proof of Lemma 5.2 in Berkes et al. [8] in order to obtain (5.80). This completes the proof. □

5.7.2 EGARCH

We appeal to Section 5.4.1, where we established the consistency of the QMLE in EGARCH. We considered a certain SRE

$$\log s_{t+1} = \Phi_t(\log s_t), \qquad t \in \mathbb{Z}, \tag{5.82}$$

in order to give a proper definition for $(\log h_t)$. Unfortunately we were not able to find a tractable expression for $\Lambda(\Phi_0^{(r)})$ when $r > 1$. This is the reason why we applied Theorem 2.6.1 to the SRE (5.82) with $r = 1$ only. The resulting contraction condition, which also guarantees the consistency of the QMLE in EGARCH, is rather awkward because one has to check it by means of stochastic simulation. In this section we encounter another unpleasant problem about EGARCH. As a matter of fact, one has to show the moment conditions N.3 (ii,iii) of Section 5.6 if one wants to establish asymptotic normality of the QMLE. This necessitates to estimate the moments of stationary solutions of general SREs $x_{t+1} = \phi_t(x_t)$, which is a nontrivial task if the random Lipschitz coefficients $(\Lambda(\phi_t))$ are dependent. Essentially one would have to find accurate moment bounds for $\prod_{t=1}^k \Lambda(\phi_t)$. We are not aware of any results solving this kind of problem, not even if $(\Lambda(\phi_t))$ fulfills mixing conditions. The random Lipschitz coefficients $(\Lambda(\Phi_t))$ appearing in EGARCH are *not* independent. As a matter of fact, we can only demonstrate the asymptotic normality of the QMLE in the subclass of EGARCH models with $\beta = 0$ because there the corresponding sequence of random Lipschitz coefficients is 1–dependent.

We follow the notation of Section 5.4.1, but consider EGARCH models with $\beta = 0$, i.e., the parameter space is of form $\Theta = \mathbb{R} \times D_E$, where $D_E = \{(\gamma, \delta)^T \in \mathbb{R}^2 \mid \gamma \in \mathbb{R}, \delta \geq |\gamma| \}$. The true parameter vector is denoted by $\boldsymbol{\theta}_0 = (\alpha_0, \gamma_0, \delta_0)^T$. We take a compact subset $K \subset \Theta$ containing $\boldsymbol{\theta}_0$ in its interior such that

$$\mathbb{E}(\log \|\lambda_0\|_K) < 0 \tag{5.83}$$

with

$$\lambda_0(\boldsymbol{\theta}) = 2^{-1} \exp(-\underline{\alpha}/2)(\gamma X_0 + \delta |X_0|),$$

where $\underline{\alpha} = \inf_{\boldsymbol{\theta} \in K}\{\alpha\}$. The distribution of Z_0 must not be concentrated in two points. Then Theorem 5.4.1 tells us that the QMLE is strongly consistent. Another implication of Theorem 5.4.1 is the fact that the SRE (5.82) has a unique stationary solution $(\log h_t)$, which is ergodic and where h_t is $\mathcal{F}_{t-1}$–measurable for every t. In what follows, we study the derivatives of $\ell_t = \log h_t + X_t^2/h_t$. Note that we cannot directly apply Theorem 5.6.1 because in EGARCH we work with a SRE for $(\log h_t)$ rather than for (h_t). The necessary modifications are given in the following. First we show that $\log h_t$ is twice continuously differentiable on a compact ball $\tilde{K}$ about $\boldsymbol{\theta}_0$ and compute moments of the derivatives. For notational ease we can assume $\tilde{K} = K$. We use the techniques of Section 5.5. In the second step, asymptotic normality will be proved along the lines of the proof of Theorem 5.6.1. Before we start, we introduce

$$\bar{\delta} = \max\Big(\sup_{\boldsymbol{\theta} \in K} (\gamma + \delta), \sup_{\boldsymbol{\theta} \in K} (\delta - \gamma) \Big)$$

and observe that $\gamma x + \delta |x| \leq \bar{\delta}\, |x|$ for all $x \in \mathbb{R}$.

Lemma 5.7.6. *Suppose the conditions imposed for this section. Then* $\log h_t$ *is continuously differentiable on* K, $(\log h_t)'$ *is* $\mathcal{F}_{t-1}$*–measurable and*

$$\|(\log \hat{h}_t)' - (\log h_t)'\|_K \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty. \tag{5.84}$$

If there is $\eta > 0$ *such that*

$$\lambda = 2^{-\eta} \exp(-\eta \underline{\alpha}/2) \bar{\delta}^{\eta} \mathbb{E}|\sigma_1 Z_0|^{\eta} < 1, \tag{5.85}$$

then $\mathbb{E}\|(\log h_t)'\|_K^{\eta} < \infty$.

Proof. Differentiation with respect to $\boldsymbol{\theta}$ of both sides of

$$\begin{aligned} \log \hat{h}_{t+1} &= g_{\boldsymbol{\theta}}(X_t, \log \hat{h}_t) \\ &= \alpha + (\gamma X_{t-1} + \delta |X_t|) \exp(-2^{-1}(\log \hat{h}_t)), \qquad t \in \mathbb{N}, \end{aligned}$$

leads to

$$(\log \hat{h}_{t+1})' = \hat{\dot{\Phi}}_t((\log \hat{h}_t)') = \hat{A}_t (\log \hat{h}_t)' + \hat{B}_t, \tag{5.86}$$

where

$$\begin{aligned} \hat{A}_t &= \frac{\partial g_{\boldsymbol{\theta}}}{\partial s}(X_t, \log \hat{h}_t) = -2^{-1}(\gamma X_t + \delta |X_t|) \exp(-2^{-1} \log \hat{h}_t), \\ \hat{B}_t &= \frac{\partial g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}(X_t, \log \hat{h}_t) = (1, X_t \exp(-2^{-1} \log \hat{h}_t), |X_t| \exp(-2^{-1} \log \hat{h}_t)). \end{aligned}$$

We mention that $\partial g_{\boldsymbol{\theta}}/\partial \boldsymbol{\theta}$ denotes the derivative with respect to the first coordinate of the function $(\boldsymbol{\theta}, s) \mapsto g_{\boldsymbol{\theta}}(x, s)$ and $\partial g_{\boldsymbol{\theta}}/\partial s$ the derivative with respect to the second coordinate. Replacing $\log \hat{h}_t$ by $\log h_t$ in $\hat{A}_t$ and $\hat{B}_t$, we obtain the corresponding stationary sequences (A_t) and (B_t). Define on $\mathbb{C}(K, \mathbb{R}^3)$ the SRE

$$q_{t+1} = \Phi_t(q_t) = A_t q_t + B_t, \qquad t \in \mathbb{Z}. \tag{5.87}$$

Since $\mathbb{E}(\log^+ |X_0|) < \infty$, $\log \hat{h}_t, \log h_t \geq \underline{\alpha}$ and $\|\log \hat{h}_t - \log h_t\|_K \xrightarrow{\text{e.a.s.}} 0$, one has

$$\Lambda(\hat{\dot{\Phi}}_t - \dot{\Phi}_t) = \|\hat{A}_t - A_t\|_K \leq 2^{-1} \bar{\delta}\, |X_t|\, \exp(2^{-1} \underline{\alpha}) \|\log \hat{h}_t - \log h_t\|_K \xrightarrow{\text{e.a.s.}} 0.$$

Moreover,

$$\|\hat{\dot{\Phi}}_t(0) - \dot{\Phi}_t(0)\|_K = \|\hat{B}_t - B_t\|_K \leq \text{const} \times |X_t|\, \|\log \hat{h}_t - \log h_t\|_K \xrightarrow{\text{e.a.s.}} 0.$$

By virtue of $\mathbb{E}(\log \Lambda(\dot{\Phi}_0)) = \mathbb{E}(\log \|A_0\|_K) \leq \mathbb{E}(\log \|\lambda_0\|_K) < 0$, the SRE (5.87) is contractive. From an application of Theorem 2.6.4 and identical arguments as used in the proof of Proposition 5.5.1, the differentiability of $\log h_t$ and the relation (5.84) follow. It remains to prove $\mathbb{E}\|(\log h_t)'\|_K^{\eta} < \infty$. Due to the linearity of the SRE (5.87), its unique stationary ergodic solution $((\log h_t)')$ has the representation

$$(\log h_t)' = \sum_{k=1}^{\infty} \Big(\prod_{i=1}^{k-1} A_{t-i} \Big) B_{t-k} \quad \text{a.s.}$$

Noticing that $\|B_t\|_K \leq c_1(1+|X_t|)$ for some constant $c_1 > 0$ and applying the triangle inequality to the latter representation, we receive

$$\|(\log h_t)'\|_K \leq c_1 \sum_{k=1}^{\infty} P_{t,k-1}(1+|X_{t-k}|), \tag{5.88}$$

where $P_{t,k}$ stands for the product of random Lipschitz coefficients $\prod_{i=1}^{k} \|A_{t-i}\|_K$ (set $P_{t,0}=1$). Note that

$$\|A_t\|_K \leq 2^{-1}\bar{\delta}|X_t| \exp(-2^{-1}\underline{\alpha}). \tag{5.89}$$

Since we consider an EGARCH model with $\beta_0 = 0$, the volatility σ_t is a function of Z_{t-1} for every t, and the decomposition

$$\prod_{i=1}^{k} X_{t-i} = Z_{t-1}\left(\prod_{i=2}^{k} \sigma_{t-i+1} Z_{t-i}\right)\sigma_{t-k}$$

consists of independent factors. Using the latter decomposition together with (5.89) and $(1+|X_{t-k}|)^\eta \leq 2^\eta(1+|X_{t-k}|^\eta)$, we obtain

$$\begin{aligned}
\mathbb{E}[P_{t,k-1}^\eta(1+|X_{t-k}|)^\eta] &= \mathbb{E}\Big((1+|X_{t-k}|)^\eta \prod_{i=1}^{k-1}\|A_{t-i}\|_K^\eta\Big)\\
&\leq 2^\eta 2^{\eta(-k+1)}\bar{\delta}^{\eta(k-1)} \exp(-2^{-1}\underline{\alpha}\eta(k-1))\\
&\quad \times\left(\mathbb{E}\Big(\prod_{i=1}^{k-1}|X_{t-i}|^\eta\Big) + \mathbb{E}\Big(\prod_{i=1}^{k}|X_{t-i}|^\eta\Big)\right)\\
&= 2^\eta 2^{\eta(-k+1)}\bar{\delta}^{\eta(k-1)} \exp(-2^{-1}\underline{\alpha}\eta(k-1))\mathbb{E}|Z_0|^\eta \mathbb{E}\sigma_0^\eta\\
&\quad \times\Big\{(\mathbb{E}|\sigma_1 Z_0|^\eta)^{k-2} + (\mathbb{E}|\sigma_1 Z_0|^\eta)^{k-1}\Big\}\\
&= c_2 \lambda^{k-2}
\end{aligned} \tag{5.90}$$

with $c_2 = \bar{\delta}^\eta \exp(2^{-1}\underline{\alpha}\eta)(\mathbb{E}|Z_0|^\eta)(\mathbb{E}\sigma_0^\eta)(1+\mathbb{E}|\sigma_1 Z_0|^\eta)$. From $\mathbb{E}|\sigma_1 Z_0|^\eta < \infty$ it is easy to see that $c_2 < \infty$. An application of the Minkowski inequality to (5.88) and incorporation of (5.90) gives

$$\begin{aligned}
\mathbb{E}\|(\log h_0)'\|_K^\eta &\leq c_1^\eta \Big\{\sum_{k=1}^{\infty}\left(\mathbb{E}[P_{t,k-1}^\eta(1+|X_{t-k}|)^\eta]\right)^{1/\max(1,\eta)}\Big\}^{\max(1,\eta)}\\
&\leq c_1^\eta \Big\{\sum_{k=1}^{\infty}\left(c_2\lambda^{k-2}\right)^{1/\max(1,\eta)}\Big\}^{\max(1,\eta)} < \infty.
\end{aligned}$$

This completes the proof. □

Lemma 5.7.7. *Suppose all the conditions of Lemma 5.7.6, in particular also the existence of $\eta > 0$ such that $\lambda < 1$. Then $\log h_t$ is twice continuously differentiable on K, $(\log h_t)''$ is $\mathcal{F}_{t-1}$–measurable and*

$$\|(\log \hat{h}_t)'' - (\log h_t)''\|_K \xrightarrow{\text{e.a.s.}} 0, \qquad t \to \infty. \tag{5.91}$$

Moreover, $\mathbb{E}\|(\log h_0)''\|_K^{\eta/2} < \infty$.

Proof. Differentiation of (5.86) with respect to $\boldsymbol{\theta}$ yields

$$(\log \hat{h}_{t+1})'' = \hat{\ddot{\Phi}}_t((\log \hat{h}_t)'') = \hat{A}_t(\log \hat{h}_t)' + \hat{C}_t, \tag{5.92}$$

where

$$\begin{aligned}\hat{C}_t &= \frac{\partial^2 g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^2}(X_t, \log \hat{h}_t) + \left(2\frac{\partial^2 g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta} \partial s}(X_t, \log \hat{h}_t)\right)^T (\log \hat{h}_t)' \\ &\quad + \frac{\partial^2 g_{\boldsymbol{\theta}}}{\partial s^2}(X_t, \log \hat{h}_t)\,((\log \hat{h}_t)')^T (\log \hat{h}_t)' \\ &= -2^{-1} \exp(-(\log \hat{h}_t)/2)\,(0, X_t, |X_t|)^T (\log \hat{h}_t)' \\ &\quad + 2^{-2}(\gamma X_t + \delta|X_t|) \exp(-(\log \hat{h}_t)/2)((\log \hat{h}_t)')^T (\log \hat{h}_t)'.\end{aligned}$$

The corresponding stationary sequence (C_t) is obtained from a substitution of $(\log \hat{h}_t, (\log \hat{h}_t)')$ by $(\log h_t, (\log h_t)')$ in $\hat{C}_t$. We define the SRE

$$q_{t+1} = \ddot{\Phi}_t(q_t) = A_t q_t + C_t, \qquad t \in \mathbb{Z},$$

on $\mathbb{C}(K, \mathbb{R}^{3\times 3})$ and mention that $\mathbb{C}(K, \mathbb{R}^{3\times 3})$ is equipped with the supremum norm (2.10) induced by the Frobenius norm. It is clear that the above SRE is contractive. It is standard to show that $\|\hat{\ddot{\Phi}}_t(0) - \ddot{\Phi}_t(0)\|_K \xrightarrow{\text{e.a.s.}} 0$ and $\Lambda(\hat{\ddot{\Phi}}_t - \ddot{\Phi}_t) \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$: use a decomposition of form (5.59) and inequality (2.11), recall that $\mathbb{E}(\log^+ |X_0|) < \infty$ and be aware that Lemma 5.7.6 implies $\mathbb{E}(\log^+ \|(\log h_t)'\|_K) < \infty$ and $\|(\log \hat{h}_t)' - (\log h_t)\|_K \xrightarrow{\text{e.a.s.}} 0$. Hence an application of Theorem 2.6.4 followed by the identical arguments as used in the proof of Proposition 5.5.2 yield (5.91). Starting point for establishing $\mathbb{E}\|(\log h_0)''\|_K^{\eta/2} < \infty$ is the a.s. representation

$$(\log h_t)'' = \sum_{k=1}^{\infty} \Big(\prod_{i=1}^{k-1} A_{t-i}\Big) C_{t-k} \quad \text{a.s.} \tag{5.93}$$

By means of inequality (2.11) one can show that there is $c_3 > 0$ so that

$$\|C_t\|_K \le c_3 |X_t|\,(\|(\log h_t)'\|_K + \|(\log h_t)'\|_K^2).$$

This estimate and an application of the triangle inequality to (5.93) gives

$$\|(\log h_t)''\|_K \le c_3 \sum_{k=1}^{\infty} P_{t,k-1}|X_{t-k}| \left(\|(\log h_{t-k})'\|_K + \|(\log h_{t-k})'\|_K^2\right)$$

$$= c_3 \sum_{k=1}^{\infty} U_{t,k} V_{t,k}, \tag{5.94}$$

where $U_{t,k} = P_{t,k-1}|Z_{t-k}|$ and $V_{t,k} = \sigma_{t-k}(\|(\log h_{t-k})'\|_K + \|(\log h_{t-k})'\|_K^2)$. Similarly to the derivation of (5.90) we can demonstrate that $\mathbb{E}U_{t,k}^{\eta} \le c_4\lambda^k$ for some $c_4 > 0$. This implies $\mathbb{E}U_{t,k}^{\eta/2} \le \sqrt{c_4}\lambda^{k/2}$. Using inequality (5.88) and the same arguments which lead to (5.90), we establish

$$\mathbb{E}\|\sigma_t^{1/2}(\log h_t)'\|_K^{\eta} < \infty \quad \text{and} \quad \mathbb{E}\|\sigma_t(\log h_t)'\|_K^{\eta} < \infty, \tag{5.95}$$

so that $\mathbb{E}V_{t,k}^{\eta/2} < \infty$ follows. Since $U_{t,k}$ and $V_{t,k}$ are independent, there is $c_5 > 0$ so that $\mathbb{E}[(U_{t,k}V_{t,k})^{\eta/2}] \le c_5\lambda^{k/2}$. This result and the Minkowski inequality applied to (5.94) show $\mathbb{E}\|(\log h_0)''\|_K^{\eta/2} < \infty$. This concludes the proof of the lemma. □

Next we treat the derivatives of $1/h_t$. Differentiation of $\exp(-\log h_t)$ leads to

$$\left(\frac{1}{h_t}\right)' = -\frac{1}{h_t}(\log h_t)', \tag{5.96}$$

$$\left(\frac{1}{h_t}\right)'' = \left(\frac{1}{h_t}\right)^2 ((\log h_t)')^T(\log h_t)' - \frac{1}{h_t}(\log h_t)''. \tag{5.97}$$

The following lemma provides limit relations and moment estimates necessary in the proof of asymptotic normality.

Lemma 5.7.8. *Under the conditions of Lemma 5.7.7,*

$$\left\|\left(\frac{1}{\hat{h}_t}\right)' - \left(\frac{1}{h_t}\right)'\right\|_K \xrightarrow{\text{e.a.s.}} 0 \quad \textit{and} \quad \left\|\left(\frac{1}{\hat{h}_t}\right)'' - \left(\frac{1}{h_t}\right)''\right\|_K \xrightarrow{\text{e.a.s.}} 0, \quad t \to \infty. \tag{5.98}$$

Moreover,

$$\mathbb{E}\left\|X_0^2\left(\frac{1}{h_0}\right)'\right\|_K^{\eta/2} < \infty \quad \textit{and} \quad \mathbb{E}\left\|X_0^2\left(\frac{1}{h_0}\right)''\right\|_K^{\eta/2} < \infty. \tag{5.99}$$

Proof. The relations (5.98) follow from the decomposition (5.59) together with arguments which are standard by now; we omit details. As a starting point for showing (5.99) we make use of the inequalities (5.89) and (5.94) to conclude $\mathbb{E}\|\sigma_0^2(\log h_0)''\|_K^{\eta/2} < \infty$. This relation together with $\mathbb{E}\|(\log h_0)'\|_K^{\eta} < \infty$, $\mathbb{E}\|(\log h_0)''\|_K^{\eta/2} < \infty$, (5.95), $h_0 \ge \exp(\underline{\alpha})$ and the facts that Z_0 is independent of $\{\log h_0, (\log h_0)', (\log h_0)''\}$ and $\mathbb{E}|Z_0|^{\eta} < \infty$ yields the desired results. □

We are now ready to demonstrate the asymptotic normality of the QMLE $\hat{\boldsymbol{\theta}}_n$ in EGARCH with $\beta = 0$.

Theorem 5.7.9. *Let (X_t) be a stationary* EGARCH *process with parameters $\boldsymbol{\theta}_0 = (\alpha_0, \gamma_0, \delta_0)^T$. Let $K \subset \mathbb{R} \times D_E$ be a compact set which contains $\boldsymbol{\theta}_0$ in its interior. Suppose that the distribution of Z_0 is not concentrated in two points and that* (5.83) *holds true. Assume that condition* (5.85) *is met with $\eta \geq 2$. Then the* QMLE $\hat{\boldsymbol{\theta}}_n$ *is strongly consistent and asymptotically normal with covariance matrix*

$$\mathbf{V}_0 = 4^{-1}\mathbb{E}(Z_0^4 - 1)\left(1 + (\gamma_0^2 + \delta_0^2 + 2\gamma_0\delta_0\mathbb{E}(Z_0|Z_0|))/4\right)$$
$$\times \left(\mathbf{U}_0 - (\mathbf{W}_0 + \mathbf{W}_0^T)/2\right)^{-1},$$

where

$$\mathbf{U}_0 = \mathbb{E}[\,(1, Z_0, |Z_0|)^T(1, Z_0, |Z_0|)\,],$$
$$\mathbf{W}_0 = (1 + \delta_0\mathbb{E}|Z_0|/2)^{-1}$$
$$\times(\delta_0\mathbb{E}|Z_0|, \delta_0\mathbb{E}(Z_0|Z_0|) + \gamma_0, \gamma_0\mathbb{E}(Z_0|Z_0|) + \delta_0)^T(1, 0, \mathbb{E}|Z_0|).$$

Proof. The relations $\mathbb{E}\|(\log h_0)'\|_K^\eta < \infty$, $\mathbb{E}\|(\log h_0)''\|_K^{\eta/2} < \infty$ and (5.99) show that $\mathbb{E}\|\ell_0''\|_K < \infty$ and $\mathbb{E}|(\log h_0)'(\boldsymbol{\theta}_0)|^2 < \infty$. The assumption $\lambda < 1$ and $\eta \geq 2$ imply $\mathbb{E}|\sigma_1 Z_0| < \infty$. Since $\boldsymbol{\theta}_0$ lies in the interior of K, there exists $\underline{\delta} > 0$ such that $\delta_0 Z_0 + \gamma_0|Z_0| \geq 2\underline{\delta}|Z_0|$. Consequently, $\mathbb{E}|Z_0 \exp(\underline{\delta}|Z_0|)| \leq \exp(-\alpha_0/2)\mathbb{E}|\sigma_1 Z_0| < \infty$, showing that the moment generating function of $|Z_0|$ is finite in a neighborhood of zero, which implies that Z_0 has finite moments of any order, in particular, $\mathbb{E}Z_0^4 < \infty$. Hence the moment conditions required by N.3 hold and all the steps in the proof of Theorem 5.6.1 are valid. We merely mention that the matrix $\mathbf{F}_0 = \mathbb{E}[\,\ell_0''(\boldsymbol{\theta}_0)\,]$ is positive definite because $(\log h_0)'(\boldsymbol{\theta}_0)\mathbf{x} = 0$ a.s. implies $(1, Z_0, |Z_0|)\mathbf{x} = 0$ a.s. for every $\mathbf{x} \in \mathbb{R}^3$, which can only be true if $\mathbf{x} = \mathbf{0}$. Note that the SRE for $(\log h_t)'$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ leads to

$$(\log h_1)'(\boldsymbol{\theta}_0) = \frac{\partial g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}(X_0, \log \sigma_0^2)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \frac{\partial g_{\boldsymbol{\theta}}}{\partial s}(X_0, \log \sigma_0^2)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}(\log h_0)'(\boldsymbol{\theta}_0)$$
$$= (1, Z_0, |Z_0|) - \frac{1}{2}(\gamma_0 Z_0 + \delta_0|Z_0|)(\log h_0)'(\boldsymbol{\theta}_0). \quad (5.100)$$

Since Z_0 and $(\log h_0)'$ are independent, taking the expectation on both sides of (5.100) gives

$$\mathbb{E}[\,(\log h_0)'(\boldsymbol{\theta}_0)\,] = (1 + \delta_0\mathbb{E}|Z_0|/2)^{-1}\ (1, 0, \mathbb{E}|Z_0|).$$

Likewise, squaring the equation (5.100) and taking the expectations on both sides yields the value of $\mathbb{E}[\,((\log h_0)'(\boldsymbol{\theta}_0))^T(\log h_0)'(\boldsymbol{\theta}_0)\,]$. Plugging the latter expression into (5.66) gives the asymptotic covariance matrix $\mathbf{V}_0$ and completes the proof of the theorem. ◻

5.8 Non–Stationarities

So far our considerations were based on the fundamental working assumption that we sample from the unique *stationary* solution to the equations (5.1) of the general heteroscedastic model. This is a strong assumption, particularly in the light of Remark 5.2.2. There we mentioned that simulating the stationary solution is in general impossible, and we provided an algorithm which generates a sequence $({}_k\tilde{\boldsymbol{\sigma}}_t^2)_{t\in\mathbb{N}}$ which approaches the stationary solution $({}_k\boldsymbol{\sigma}_t^2)_{t\in\mathbb{N}}$ of (5.5) with an error decaying to zero exponentially fast as $t \to \infty$:

1. Take an initial value $\varsigma_0^2 \in [0,\infty)^k$, set ${}_k\tilde{\boldsymbol{\sigma}}_0^2 = \varsigma_0^2$, and generate ${}_k\tilde{\boldsymbol{\sigma}}_t^2$ according to (5.5).
2. Set $(\tilde{\sigma}_t^2, \tilde{X}_t) = \left({}_k\tilde{\boldsymbol{\sigma}}_{t,1}^2\right)^{1/2}(1, Z_t), \qquad t \in \mathbb{N}.$

Notice that $\varsigma_0 \in [0,\infty)^q$ is an arbitrarily chosen initial value. Correspondingly we define $\tilde{X}_t = \tilde{\sigma}_t Z_t$ and observe that $|\tilde{X}_t^2 - X_t^2| = |Z_t|\,|\tilde{\sigma}_t - \sigma_t| \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$ by virtue of Lemma 2.5.4. To the best of our knowledge, simulation studies have been based on random samples of $(\tilde{X}_t, t = 0,\ldots,n)$ rather than of $(X_t, t = 0,\ldots n)$, see e.g. Lumsdaine [89]. The effects of this inherent non–stationarity due to some initialization error have hardly ever been addressed. Although simulation studies are most often carried out for determining the small sample behavior of estimators, from a theoretical point of view a minimal requirement to validate the simulation approach would be the asymptotic equivalence of the QMLE applied to $(\tilde{X}_t)_{t\in\mathbb{N}}$ and to $(X_t)_{t\in\mathbb{N}}$. Let

$$\tilde{L}_n = -\frac{1}{2}\sum_{t=1}^{n}(X_t^2/\tilde{h}_t + \log\tilde{h}_t)$$

denote the likelihood based on the observations $(\tilde{X}_t, t = 0,\ldots,n)$ (here we set $\tilde{x}_t = 0$ for $t < 0$ such that $\tilde{\mathbf{h}}_0$ can be properly defined). To establish the asymptotic equivalence of $\hat{\boldsymbol{\theta}}_n$ and of the maximizer of $\tilde{L}_n$, we merely need to show

$$n^{-1}\,\|\tilde{L}_n - \hat{L}_n\|_K \xrightarrow{\text{e.a.s.}} 0, \qquad n \to \infty \tag{5.101}$$

and

$$n^{-1/2}\,\|\tilde{L}_n' - \hat{L}_n'\|_K \xrightarrow{\text{e.a.s.}} 0, \qquad n \to \infty. \tag{5.102}$$

One can impose conditions on the functions $g_{\boldsymbol{\theta}}$, which imply (5.101) and (5.102), but since everything is in line with the ideas already presented in this chapter, we omit details and merely mention that the QMLE in nonstationary AGARCH(p,q) has limit behavior identical to when applied to stationary data.

5.9 Fitting AGARCH(1,1) to the NYSE Composite Data

As an illustration of the QMLE, we fit the AGARCH$(1,1)$ model to the NYSE Composite log–returns, which were already considered in Chapter 3. For obtaining a fit, we use the routine `garch`, of the commercial software package `S+Finmetrics` [69]; for a detailed description of the features of `S+Finmetrics` we refer to Zivot and Wang [136]. The estimates

$$\hat{\alpha}_0 = 1.492 \times 10^{-6},\ \hat{\alpha}_1 = 0.06706,\ \hat{\beta}_1 = 0.9079,\ \hat{\gamma}_1 = 0.3594$$

have standard errors

$$\widehat{\mathrm{se}}(\hat{\alpha}_0) = 4.910 \times 10^{-7},\ \widehat{\mathrm{se}}(\hat{\alpha}_1) = 0.0144,\ \widehat{\mathrm{se}}(\hat{\beta}_1) = 0.0194,\ \widehat{\mathrm{se}}(\hat{\gamma}_1) = 0.0415.$$

The standard errors are based on the normal approximation provided by Theorem 5.7.1, where the asymptotic covariance matrix V_0 is replaced by its consistent estimator $\hat{V}_0^{(n)}$, which was defined in Remark 5.6.2. Unfortunately, the documentation to `S+Finmetrics` does not give any information about the chosen initialization $\hat{h}_0(\boldsymbol{\theta}) = \varsigma_0^2$. Recall that we base the parameter estimation on NYSE Composite data of the time period January 3, 1966 – January 28, 2003 ($n = 9328$). Over such a long period of time, the stationarity hypothesis might be unrealistic; see Mikosch and Stărică [101, 101] for a critical appraisal of the working hypothesis of stationarity.

In Figure 5.1 we graph the NYSE Composite log–returns together with the estimated AGARCH$(1,1)$ volatilities. Also compare with the ad–hoc EWMA method used to produce Figure 3.4. In Figure 5.2 we provide a simple graphical analysis of the residuals $\hat{Z}_t^{(n)} = X_t/(\hat{h}_t(\hat{\boldsymbol{\theta}}_n))^{1/2}$ in order to assess the goodness–of–fit. As predicted by the asymptotic theory of the QMLE in AGARCH$(1,1)$, the sample autocorrelations of $|\hat{Z}_t^{(n)}|$ are close to zero. The estimated absolute residuals do not reveal any linear dependence; this is clearly in contrast to Figure 3.3, which depicts that the sample autocorrelations of the absolute log–returns do not vanish. Moreover, the QQ–plot of the residuals against standard normal indicates that the distribution of the innovations Z_t is heavy–tailed (with respect to the Gaussian family) and skewed. The numbers $\hat{\gamma}_1 = 0.3594$ and $\widehat{\mathrm{se}}(\hat{\gamma}_1) = 0.0415$ indicate that there are significant leverage effects since an (asymptotic) test of the hypothesis $\gamma_1^\circ = 0$ would be rejected at any reasonable level of significance; compare with Figure 3.7, where the leverage effects were identified graphically.

5.10 A Simulation Study

In order to investigate the small, moderate and large sample behavior of the QMLE in AGARCH$(1,1)$, we conduct a simulation study. We consider three models, which all have the parameters estimated from the NYSE Composite log–return data in the previous section:

$$\alpha_0^\circ = 1.492 \times 10^{-6},\ \alpha_1^\circ = 0.06706,\ \beta_1^\circ = 0.9079,\ \gamma_1^\circ = 0.3594. \qquad (5.103)$$

We assume that the innovations Z_t have Student t_ν density with $\nu = \infty, 9, 5$ (the value $\nu = \infty$ corresponds to standard normal innovations). Note that

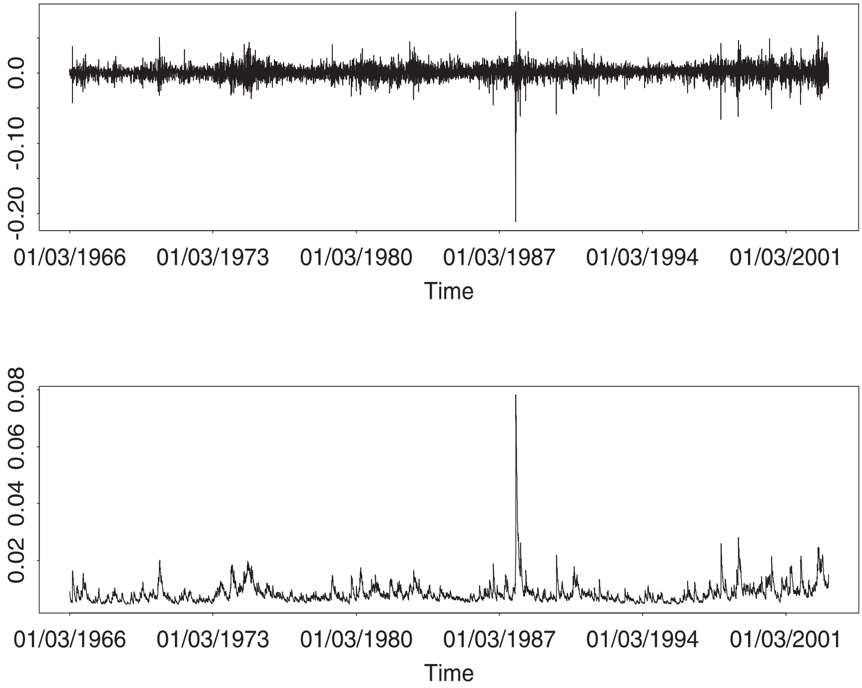

Fig. 5.1. NYSE Composite log–returns (top) and volatilities estimated through a fit of the AGARCH(1, 1) model (bottom).

we have to standardize the t_ν densities such that the variance is 1; see (6.12). Observe that $\mathbb{E}Z_0^4 < \infty$, and hence the statements of Theorem 5.7.1 apply. By using the `S+Finmetrics` [69] function `simulate.garch` we generate for each model time series of lengths $n = 250$, 500, 1000, 5000 and 10000, to which the function `garch` is applied in a second step in order to obtain parameter estimates. We repeat this procedure 500 times and hence get 500 independent replicates of $\hat{\boldsymbol{\theta}}_n$ for each model and sample size $n = 250$, 500, 1000, 5000, 10000.

In what follows we analyze the distributions of the vectors $\hat{\boldsymbol{\theta}}_n$ by compiling boxplots and tables for each estimated parameter. The outcome is slightly disappointing. A glance at the figures below shows that in the three models considered, the QMLE is not very reliable for small and moderate sample sizes ($n = 250, 500, 1000$). Particularly the distribution of $\hat{\alpha}_0$ has a huge "relative variation", even for $n = 5000$ and $n = 10000$. For all sample sizes the parameter α_0 is systematically overestimated, with a bias which is large in comparison to the true value α_0°. This feature is rather disturbing when contrasted to common best–practice risk management methods, which often base estimation on one year of daily log–returns, i.e., $n \approx 250$, see e.g. McNeil and Frey [95]. In this context it is also worth mentioning that in each of the three

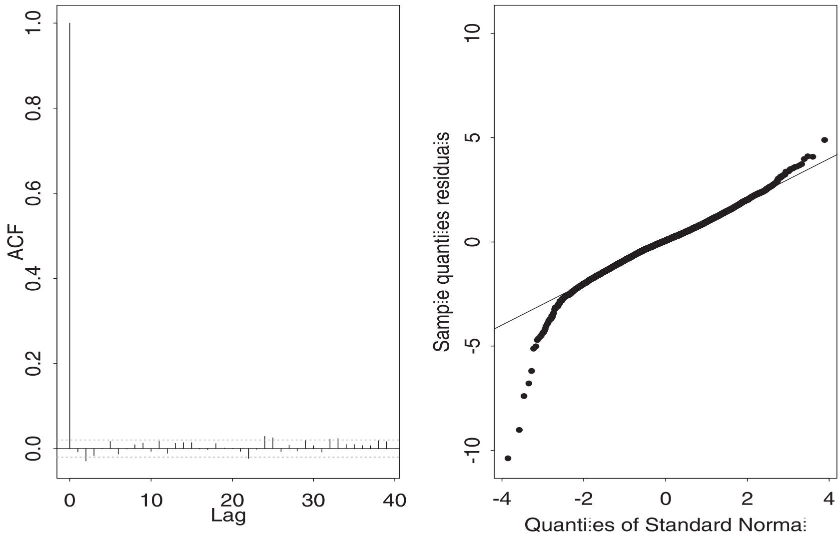

Fig. 5.2. NYSE Composite residuals obtained from a fit of AGARCH(1, 1). *Top*: Sample autocorrelations of absolute residuals. *Bottom*: QQ–plot of residuals against standard normal.

models the simulation runs sometimes yielded values $\hat{\beta}_1$ close or equal to zero when the sample size is $n = 250$. We conjecture that this is caused by the fact that α_1° is rather close to the plane $\alpha_1 = 0$, on which the parameters are nonidentifiable; cf. Hannan and Deistler [65] for a thorough treatment of identifiability issues in linear models. Moreover it seems that $\hat{\beta}_1$ systematically underestimates β_1°, with a bias which is of the order $\hat{\text{se}}(\hat{\beta}_1)$. In contrast, $\hat{\alpha}_1$ has a positive bias. It seems difficult to estimate the leverage parameter γ_1 since the distribution of $\hat{\gamma}_1$ is rather scattered, even for sample sizes $n = 10000$. As a general observation we mention that the finite sample distributions of the QMLE are the more scattered the heavier the tails of the innovations Z_t, i.e., the smaller ν.

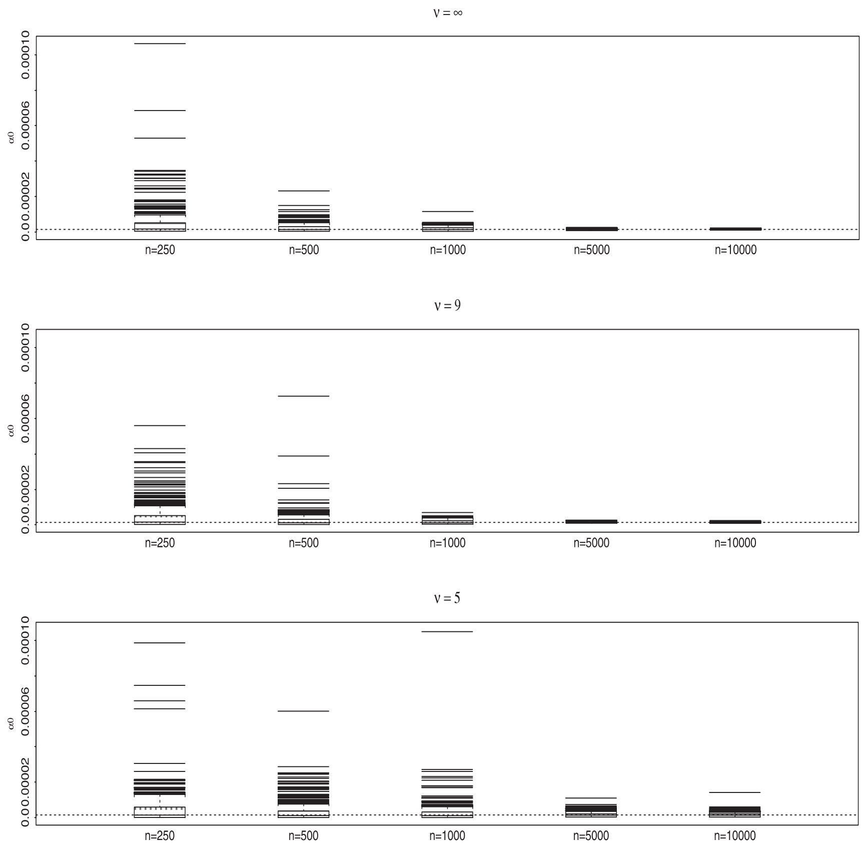

Fig. 5.3. Boxplots of independent realizations of the QMLE of $\hat{\alpha}_0$ in the AGARCH(1,1) process with Student t_ν innovations and parameters (5.103). Various sample sizes n and degrees of freedom ν are compared. The dotted horizontal lines represent the true value α_0°.

ν	n	mean	median	std. dev.	std. err.
∞	250	5.632×10^{-6}	2.834×10^{-6}	1.015×10^{-5}	5.191×10^{-6}
	500	2.422×10^{-6}	1.916×10^{-6}	2.076×10^{-6}	1.649×10^{-6}
	1000	1.849×10^{-6}	1.651×10^{-6}	9.522×10^{-7}	7.921×10^{-7}
	5000	1.559×10^{-6}	1.551×10^{-6}	2.846×10^{-7}	2.836×10^{-7}
	10000	1.527×10^{-6}	1.504×10^{-6}	1.909×10^{-7}	1.955×10^{-7}
9	250	4.758×10^{-6}	2.714×10^{-6}	6.512×10^{-6}	5.123×10^{-6}
	500	2.764×10^{-6}	1.914×10^{-6}	4.252×10^{-6}	4.986×10^{-6}
	1000	1.829×10^{-6}	1.631×10^{-6}	8.851×10^{-7}	8.683×10^{-7}
	5000	1.579×10^{-6}	1.555×10^{-6}	3.169×10^{-7}	3.212×10^{-7}
	10000	1.558×10^{-6}	1.540×10^{-6}	2.211×10^{-7}	2.226×10^{-7}
5	250	5.332×10^{-6}	3.159×10^{-6}	7.736×10^{-6}	3.946×10^{-5}
	500	2.650×10^{-6}	2.010×10^{-6}	2.655×10^{-6}	2.659×10^{-6}
	1000	2.066×10^{-6}	1.750×10^{-6}	1.506×10^{-6}	1.212×10^{-6}
	5000	1.631×10^{-6}	1.608×10^{-6}	4.090×10^{-7}	4.063×10^{-7}
	10000	1.561×10^{-6}	1.542×10^{-6}	2.987×10^{-7}	2.838×10^{-7}

Table 5.1. Key figures of the finite sample distribution of $\hat{\alpha}_0$ for various n and ν. The column std. err. contains the sample means of the standard errors $\widehat{\text{se}}(\hat{\alpha}_0)$, which are based on the normal approximation to $\hat{\alpha}_0$ and estimated with the method of Remark 5.6.2. The column std. dev. consists of the sample standard deviations of the replicates of $\hat{\alpha}_0$.

ν	n	mean	median	std. dev.	std. err.
∞	250	0.09312	0.08769	0.040860	0.058050
	500	0.07147	0.06865	0.025900	0.032570
	1000	0.06825	0.06732	0.018580	0.020940
	5000	0.06738	0.06722	0.008154	0.008441
	10000	0.06782	0.06775	0.005547	0.005866
9	250	0.06582	0.05872	0.046040	0.046380
	500	0.06365	0.05982	0.032060	0.032670
	1000	0.06355	0.06317	0.023940	0.020160
	5000	0.06745	0.06680	0.009526	0.009469
	10000	0.06728	0.06735	0.006862	0.006655
5	250	0.07467	0.05993	0.062800	0.050680
	500	0.06692	0.05990	0.045120	0.047510
	1000	0.06336	0.06271	0.027030	0.023260
	5000	0.06751	0.06711	0.013180	0.011860
	10000	0.06831	0.06744	0.013030	0.009313

Table 5.2. Key figures of the finite sample distribution of $\hat{\alpha}_1$ for various n and ν. The column std. err. contains the sample means of the standard errors $\widehat{\text{se}}(\hat{\alpha}_1)$, which are based on the normal approximation to $\hat{\alpha}_1$ and estimated with the method of Remark 5.6.2. The column std. dev. consists of the sample standard deviations of the replicates of $\hat{\alpha}_1$.

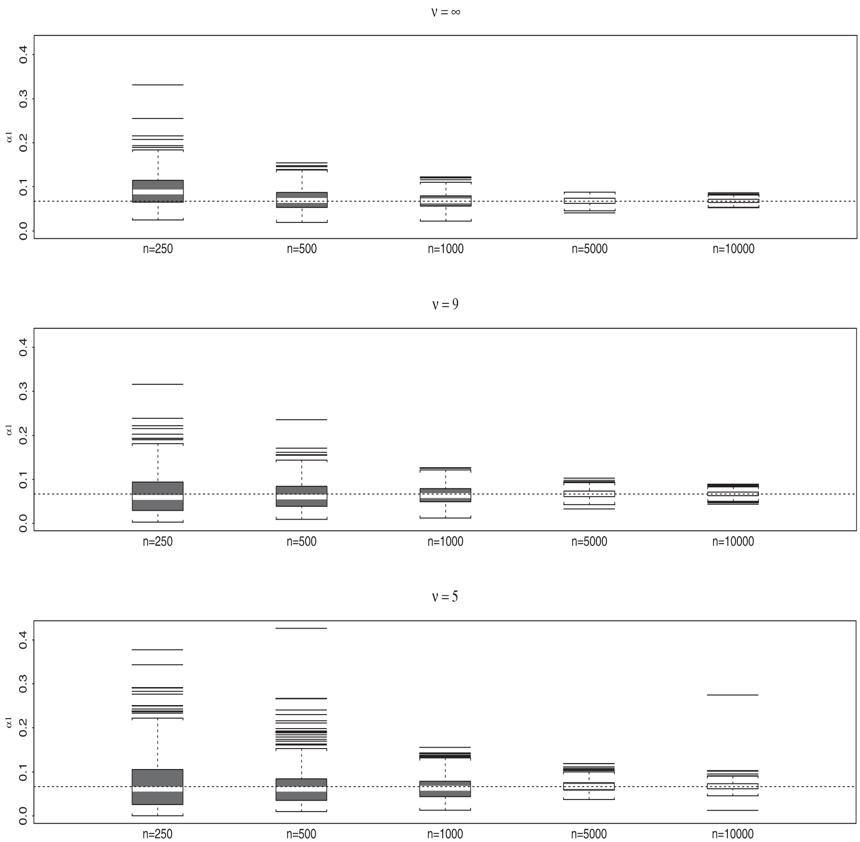

Fig. 5.4. Boxplots of independent realizations of the QMLE of $\hat{\alpha}_1$ in the AGARCH$(1,1)$ process with parameters (5.103) and Student t_ν innovations and parameters (5.103). Various sample sizes n and degrees of freedom ν are compared. The dotted horizontal lines represent the true value α_1°.

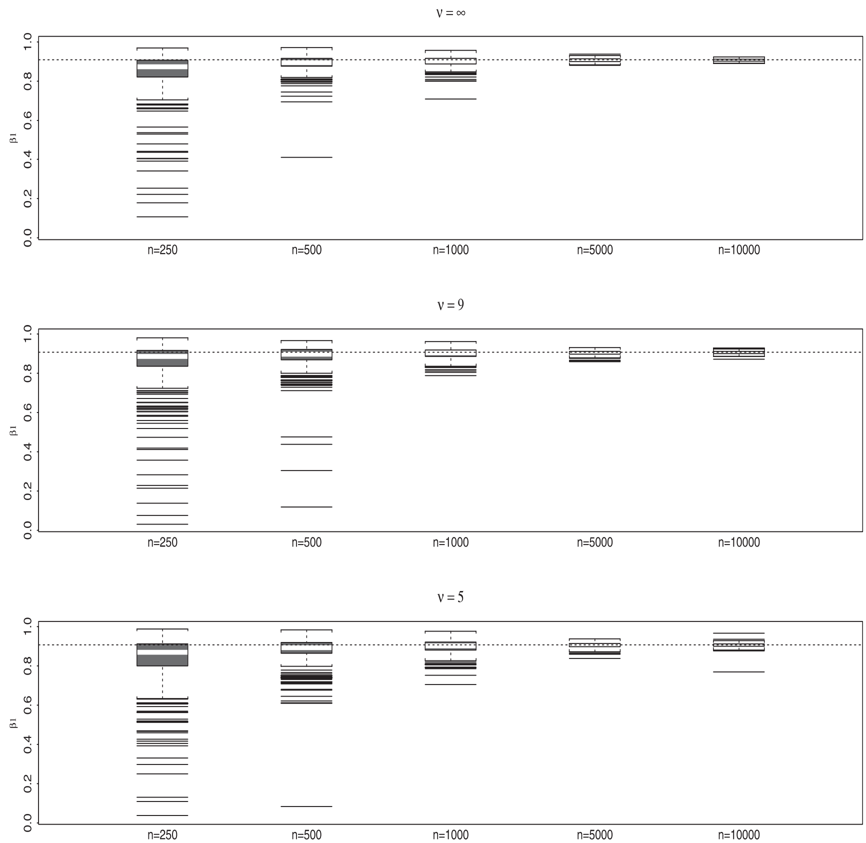

Fig. 5.5. Boxplots of independent realizations of the QMLE of $\hat{\beta}_1$ in the AGARCH(1,1) process with Student t_ν innovations and parameters (5.103). Various sample sizes n and degrees of freedom ν are compared. The dotted horizontal lines represent the true value β_1°.

ν	n	mean	median	std. dev.	std. err.
∞	250	0.8327	0.8721	0.129800	0.098550
	500	0.8900	0.8946	0.045310	0.041910
	1000	0.9004	0.9026	0.024770	0.023240
	5000	0.9062	0.9062	0.008620	0.009049
	10000	0.9064	0.9067	0.005987	0.006318
9	250	0.8539	0.8860	0.123600	0.108800
	500	0.8877	0.8994	0.067460	0.098850
	1000	0.9033	0.9061	0.025750	0.025740
	5000	0.9059	0.9060	0.010400	0.010430
	10000	0.9065	0.9069	0.007576	0.007283
5	250	0.8305	0.8699	0.133700	0.172900
	500	0.8827	0.8929	0.066430	0.072770
	1000	0.8985	0.9016	0.032490	0.031850
	5000	0.9046	0.9051	0.013480	0.013470
	10000	0.9056	0.9058	0.011390	0.009988

Table 5.3. Key figures of the finite sample distribution of $\hat{\beta}_1$ for various n and ν. The column std. err. contains the sample means of the standard errors $\widehat{\mathrm{se}}(\hat{\beta}_1)$, which are based on the normal approximation to $\hat{\beta}_1$ and estimated with the method of Remark 5.6.2. The column std. dev. consists of the sample standard deviations of the replicates of $\hat{\beta}_1$.

ν	n	mean	median	std. dev.	std. err.
∞	250	0.2728	0.2683	0.19420	0.33480
	500	0.3445	0.3312	0.16780	0.27450
	1000	0.3774	0.3632	0.13410	0.18200
	5000	0.3666	0.3597	0.06143	0.06417
	10000	0.3580	0.3576	0.04013	0.04306
5	250	0.5074	0.4405	0.42360	0.33940
	500	0.4801	0.3966	0.31650	0.27160
	1000	0.4396	0.3766	0.24140	0.18690
	5000	0.3687	0.3648	0.07185	0.07141
	10000	0.3648	0.3603	0.05060	0.04949
9	250	0.4975	0.5012	0.45940	0.34600
	500	0.5051	0.4352	0.34510	0.29530
	1000	0.4549	0.4007	0.26020	0.21310
	5000	0.3706	0.3599	0.09792	0.08781
	10000	0.3613	0.3556	0.07228	0.06084

Table 5.4. Key figures of the finite sample distribution of $\hat{\gamma}_1$ for various n and ν. The column std. err. contains the sample means of the standard errors $\widehat{\mathrm{se}}(\hat{\gamma}_1)$, which are based on the normal approximation to $\hat{\gamma}_1$ and estimated with the method of Remark 5.6.2. The column std. dev. consists of the sample standard deviations of the replicates of $\hat{\gamma}_1$.

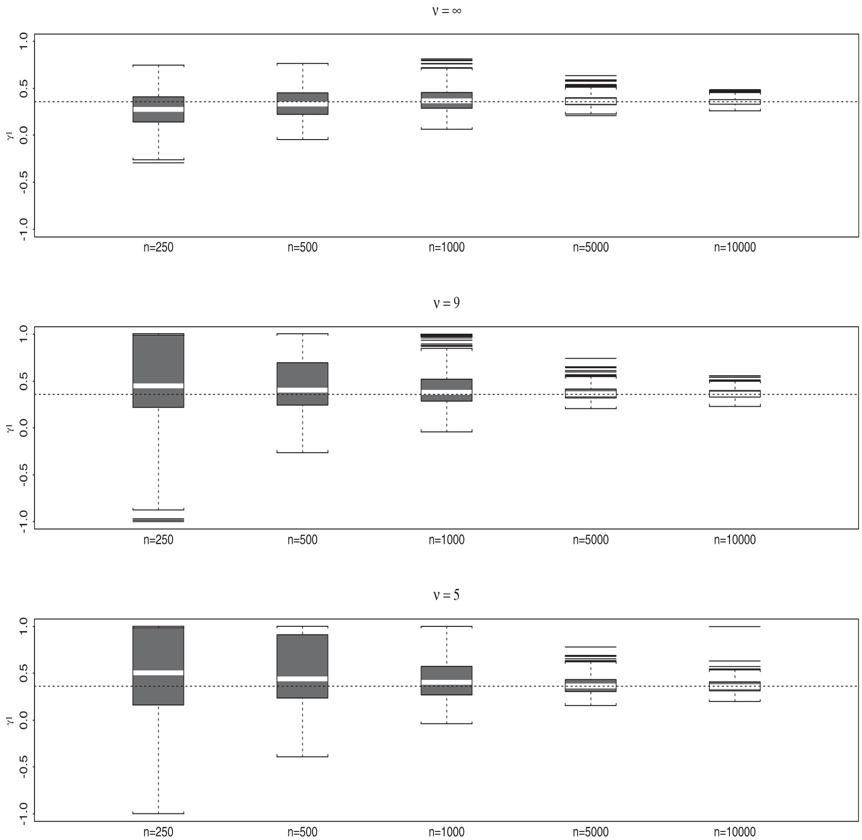

Fig. 5.6. Boxplots of independent realizations of the QMLE of $\hat{\gamma}_1$ in the AGARCH$(1,1)$ process with Student t_ν innovations and parameters (5.103). Various sample sizes n and degrees of freedom ν are compared. The dotted horizontal lines represent the true value γ_1°.

6

Maximum Likelihood Estimation in Conditionally Heteroscedastic Time Series Models

The GARCH(p,q) model with Student t_ν innovations has attracted some attention in the literature and is considered as a realistic model for real–life log–return data. The fact that the `S+FinMetrics` [69] software package contains a routine which fits this model to data indicates that it is also popular among practitioners.

It is well accepted that the estimated innovations of real–life log–return data often exhibit heavy tails, see e.g Bollerslev [19], Baillie and Bollerslev [4] or McNeil and Frey [95] for convincing empirical evidence. One possible way for capturing this feature is to consider GARCH models with Student t_ν innovations, as proposed in Bollerslev [19] and Baillie and Bollerslev [4]. Concerning the maximum likelihood estimator (MLE) of GARCH(p,q), Engle and González–Rivera [47] and Drost and González–Rivera [41] discuss the efficiency gain with respect to the quasi maximum likelihood estimator (QMLE). Since both articles Engle and González–Rivera [47] and Drost and González–Rivera [41] do not contain rigorous proofs for the consistency and asymptotic normality of the MLE, many of the authors' conclusions are somewhat heuristic.

In this chapter we analyze the maximum likelihood estimator (MLE) in the general conditionally heteroscedastic model (5.1) of Chapter 5. We assume that the iid innovations Z_t have a Lebesgue density $k_{\boldsymbol{\nu}}$, which belongs to a known and well–specified class of densities $\mathcal{D}$, which is parametrized by $\boldsymbol{\nu}$. The (unknown) finite dimensional nuisance parameter $\boldsymbol{\nu}$ is jointly estimated with $\boldsymbol{\theta}$. We establish consistency and asymptotic normality of the resulting MLE. In view of Theorems 5.3.1 and 5.6.1, the analysis of the MLE does not really pose any additional problems as compared to the QMLE apart from notation and extra regularity assumptions. Because of the practical relevance, we will state precise conditions for consistency and asymptotic normality of the MLE in (A)GARCH(p,q), and as an important example we cover the special case of Student t_ν innovations. Moreover we discuss the problems related to misspecification of the class $\mathcal{D}$. In general, misspecification of the innovations distribution leads to inconsistent maximum likelihood estimates.

To lay down the framework, let us summarize as follows. We consider the general conditionally heteroscedastic time series model of Chapter 5, i.e.,

$$\begin{cases} X_t = \sigma_t Z_t, \\ \sigma_t^2 = g_{\boldsymbol{\theta}}(X_{t-1}, \ldots, X_{t-p}, \sigma_{t-1}^2, \ldots, \sigma_{t-q}^2), \end{cases} \qquad t \in \mathbb{Z}, \tag{6.1}$$

where the volatility process (σ_t) is nonnegative and $\{g_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in K\}$ denotes a parametric class of nonnegative functions on $\mathbb{R}^p \times [0, \infty)^q$ fulfilling conditions to be specified later. Concerning the innovations (Z_t), we additionally suppose

$$Z_t \ \text{ iid } \sim k_{\boldsymbol{\nu}}(x)\,\mathrm{d}x, \tag{6.2}$$

where $\{k_{\boldsymbol{\nu}} \mid \boldsymbol{\nu} \in V\}$ is a parametric class of Lebesgue densities on $\mathbb{R}$ with

$$\int_{-\infty}^{\infty} x\, k_{\boldsymbol{\nu}}(x)\,\mathrm{d}x = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} x^2\, k_{\boldsymbol{\nu}}(x)\,\mathrm{d}x = 1 \qquad \text{for all } \boldsymbol{\nu}.$$

We assume that $K \subset \mathbb{R}^d$ and $V \subset \mathbb{R}^{d'}$ are compact sets and use the vector notation

$$\boldsymbol{\tau} = \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\nu} \end{pmatrix}$$

for the parameters. In order to define the maximum likelihood estimator of $\boldsymbol{\tau}$, we observe that the approximate conditional log–likelihood is equal to

$$\hat{L}_n(\boldsymbol{\tau}) = \sum_{t=1}^{n} \log \hat{f}_t(\boldsymbol{\tau}), \tag{6.3}$$

where

$$\hat{f}_t(\boldsymbol{\tau}) = \frac{1}{(\hat{h}_t(\boldsymbol{\theta}))^{1/2}}\, k_{\boldsymbol{\nu}}\left(\frac{X_t}{(\hat{h}_t(\boldsymbol{\theta}))^{1/2}} \right)$$

with the random function $\hat{h}_t(\boldsymbol{\theta})$ as in Chapter 5. With the notation $\hat{\mathbf{h}}_t = (\hat{h}_t, \ldots, \hat{h}_{t-q+1})^T$ this means that

$$\hat{\mathbf{h}}_t = \begin{cases} \varsigma_0^2, & t = 0, \\ \Phi_{t-1}(\hat{\mathbf{h}}_{t-1}), & t \geq 1, \end{cases}$$

where the random maps $\Phi_t : \mathbb{C}(K, [0,\infty)^q) \to \mathbb{C}(K, [0,\infty)^q)$ are given by

$$[\Phi_t(\mathbf{s})](\boldsymbol{\theta}) = (g_{\boldsymbol{\theta}}(\mathbf{X}_t, \mathbf{s}(\boldsymbol{\theta})), s_1(\boldsymbol{\theta}), \ldots, s_{q-1}(\boldsymbol{\theta}))^T, \qquad t \in \mathbb{Z},$$

for every $\mathbf{s} \in \mathbb{C}(K, [0,\infty)^q)$. We remind that $\mathbf{X}_t = (X_t, \ldots, X_{t-q+1})^T$, that ς_0^2 is an arbitrary initial value and that $X_0, \ldots, X_{-p+1}$ are assumed to be observed. We now define the MLE of $\boldsymbol{\tau}$ as follows:

$$\hat{\boldsymbol{\tau}}_n = \operatorname{argmax}_{\boldsymbol{\tau} \in K \times V} \hat{L}_n(\boldsymbol{\tau}). \tag{6.4}$$

Under the conditions of Proposition 5.2.12, the sequence $(\hat{\mathbf{h}}_t)_{t\in\mathbb{N}}$ can be approximated by a stationary ergodic sequence $(\mathbf{h}_t)_{t\in\mathbb{N}}$ such that the error $\|\hat{\mathbf{h}}_t - \mathbf{h}_t\|_K \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$. The sequence $(\mathbf{h}_t)$ is characterized as the unique stationary solution to the stochastic recurrence equation (SRE)

$$\mathbf{s}_{t+1} = \Phi_t(\mathbf{s}_t), \qquad t \in \mathbb{Z},$$

on $\mathbb{C}(K, [0,\infty)^q)$. Analogously to Chapter 5, we introduce

$$L_n(\boldsymbol{\tau}) = \sum_{t=1}^{n} \log f_t(\boldsymbol{\tau}), \tag{6.5}$$

where

$$f_t(\boldsymbol{\tau}) = \frac{1}{(h_t(\boldsymbol{\theta}))^{1/2}}\, k_{\boldsymbol{\nu}}\left(\frac{X_t}{(h_t(\boldsymbol{\theta}))^{1/2}}\right).$$

The summands in (6.5) are stationary and $n^{-1}L_n$ approximates $n^{-1}\hat{L}_n$ with an error decaying to zero as $n \to \infty$. Observe that in Chapter 5 the functions L_n and $\hat{L}_n$ signified conditional Gaussian log–likelihoods, whereas in this chapter L_n and $\hat{L}_n$ denote conditional log–likelihoods in model (6.1)–(6.2).

6.1 Consistency of the MLE

We now list several regularity assumptions for the class

$$\mathcal{D} = \{k_{\boldsymbol{\nu}} \,|\, \boldsymbol{\nu} \in V\}$$

of densities, which will be needed for establishing the strong consistency of the MLE $\hat{\boldsymbol{\tau}}_n$:

M.1 The densities $k_{\boldsymbol{\nu}}$ are strictly positive, i.e., for all $\boldsymbol{\nu} \in V$ and $x \in \mathbb{R}$, one has $k_{\boldsymbol{\nu}}(x) > 0$.

M.2 The map $\mathbb{R} \times V \to (0,\infty)$: $\quad (x, \boldsymbol{\nu}) \mapsto k_{\boldsymbol{\nu}}(x)$ is continuous.

M.3 From $k_{\boldsymbol{\nu}}(x) = k_{\boldsymbol{\nu}'}(x)$ for all $x \in \mathbb{R}$ it follows that $\boldsymbol{\nu} = \boldsymbol{\nu}'$.

M.1 and M.2 are technical assumptions and M.3 guarantees the identifiability of the parameter $\boldsymbol{\nu}$.

6.1.1 Main Result

The subject of the following theorem is the strong consistency of $\hat{\boldsymbol{\tau}}_n$.

Theorem 6.1.1. *Let* (X_t) *be a stationary process of form* (6.1)–(6.2) *with true parameter vector* $\boldsymbol{\tau}_0 = \binom{\boldsymbol{\theta}_0}{\boldsymbol{\nu}_0}$. *Let the compact set* K *be such that the conditions* C.1 – C.4 *of Section 5.3 hold. Assume additionally:*

1. *The class* $\mathcal{D}$ *meets the conditions* M.1 – M.3 *and* $\boldsymbol{\nu}_0 \in V$.
2. *The true innovations density* $k_{\boldsymbol{\nu}_0}$ *fulfills*
$$\mathbb{E}|\log k_{\boldsymbol{\nu}_0}(Z_0)| < \infty. \tag{6.6}$$
3. $\mathcal{D}$ *is such that*
$$n^{-1}\|\hat{L}_n - L_n\|_{K\times V} \xrightarrow{\text{a.s.}} 0, \qquad n \to \infty. \tag{6.7}$$

Then the MLE *defined by* (6.4) *is strongly consistent, i.e.,* $\hat{\boldsymbol{\tau}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\tau}_0$.

Proof. One can follow the lines of proof of Theorem 5.3.1. The only step which requires additional arguments is to establish that the objective function $L(\boldsymbol{\tau}) = \mathbb{E}[\log f_0(\boldsymbol{\tau})]$ is indeed uniquely maximized at $\boldsymbol{\tau} = \boldsymbol{\tau}_0$.

Observe that

$$L(\boldsymbol{\tau}_0) = \mathbb{E}[\log f_0(\boldsymbol{\tau}_0)] = -\frac{1}{2}\mathbb{E}(\log \sigma_0^2) + \mathbb{E}[\log k_{\boldsymbol{\nu}_0}(Z_0)]$$

is finite because of $\mathbb{E}|\log \sigma_0^2| < \infty$ and (6.6). Since $\log x \le x - 1$ for all $x > 0$ with equality if and only if $x = 1$, we obtain

$$L(\boldsymbol{\tau}) - L(\boldsymbol{\tau}_0) = \mathbb{E}\left(\log \frac{f_0(\boldsymbol{\tau})}{f_0(\boldsymbol{\tau}_0)}\right) \le \mathbb{E}\left(\frac{f_0(\boldsymbol{\tau})}{f_0(\boldsymbol{\tau}_0)}\right) - 1 \tag{6.8}$$

with equality if and only if $f_0(\boldsymbol{\tau}) = f_0(\boldsymbol{\tau}_0)$ a.s. We now evaluate the expectation $\mathbb{E}[f_0(\boldsymbol{\tau})/f_0(\boldsymbol{\tau}_0)]$. Set

$$r(\boldsymbol{\theta}) = \frac{\sigma_0}{\sqrt{h_0(\boldsymbol{\theta})}}\,.$$

By recalling that $X_0 = \sigma_0 Z_0 = \sqrt{h_0(\boldsymbol{\theta}_0)}Z_0$ and that $r(\boldsymbol{\theta})$ is independent of $Z_0 \sim k_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x$, one concludes that

$$\mathbb{E}\left(\frac{f_0(\boldsymbol{\tau})}{f_0(\boldsymbol{\tau}_0)}\right) = \mathbb{E}\Big(\mathbb{E}\Big[\frac{r(\boldsymbol{\theta})\,k_{\boldsymbol{\nu}}(r(\boldsymbol{\theta})Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)}\,\Big|\,r(\boldsymbol{\theta})\Big]\Big) = \mathbb{E}(1) = 1.$$

Together with inequality (6.8) this implies $L(\boldsymbol{\tau}) \le L(\boldsymbol{\tau}_0)$ with equality if and only if $f_0(\boldsymbol{\tau}) = f_0(\boldsymbol{\tau}_0)$ a.s.

Since $f_0(\boldsymbol{\tau})$ contains the random scale $\sqrt{h_0(\boldsymbol{\theta})}$, a conditioning argument is needed for proving that $f_0(\boldsymbol{\tau}) = f_0(\boldsymbol{\tau}_0)$ a.s. implies $\boldsymbol{\tau} = \boldsymbol{\tau}_0$. We observe that $f_0(\boldsymbol{\tau}) = f_0(\boldsymbol{\tau}_0)$ a.s. is equivalent to

$$k_{\boldsymbol{\nu}_0}(Z_0) = r(\boldsymbol{\theta})\,k_{\boldsymbol{\nu}}(r(\boldsymbol{\theta})Z_0) \qquad \text{a.s.} \tag{6.9}$$

As an auxiliary tool for showing that (6.9) entails $\boldsymbol{\tau} = \boldsymbol{\tau}_0$, the following observation is useful.

Lemma 6.1.2. *Let $a > 0$ be a constant. Then for all $\boldsymbol{\nu} \in V$:*

$$a \neq 1 \implies \mathbb{P}[k_{\boldsymbol{\nu}_0}(Z_0) \neq a k_{\boldsymbol{\nu}}(aZ_0)] > 0.$$

Proof. Indeed, because Z_0 has Lebesgue density $k_{\boldsymbol{\nu}_0}$, which is positive and continuous on $\mathbb{R}$, the relation $\mathbb{P}[k_{\boldsymbol{\nu}_0}(Z_0) \neq a k_{\boldsymbol{\nu}}(aZ_0)] = 0$ implies $k_{\boldsymbol{\nu}_0}(x) = a k_{\boldsymbol{\nu}}(ax)$ for all $x \in \mathbb{R}$. Hence

$$1 = \int_{-\infty}^{\infty} x^2 k_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} x^2 a k_{\boldsymbol{\nu}}(ax)\,\mathrm{d}x = 1/a^2,$$

and therefore $a = 1$ follows. Thus $\mathbb{P}[k_{\boldsymbol{\nu}_0}(Z_0) \neq a k_{\boldsymbol{\nu}}(aZ_0)] = 0$ implies $a = 1$, which is the counterposition of the implication to be proved. Thus we have completed the proof of the lemma. □

We now show that (6.9) implies $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\boldsymbol{\nu} = \boldsymbol{\nu}_0$. First, suppose by contradiction $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Then, since $\sigma_0^2 = h_0(\boldsymbol{\theta})$ a.s. if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, as assumed in Condition C.4, we have $\mathbb{P}[r(\boldsymbol{\theta}) \neq 1] > 0$. From Lemma 6.1.2 we conclude that

$$\mathbb{P}[\, k_{\boldsymbol{\nu}_0}(Z_0) \neq r(\boldsymbol{\theta}) k_{\boldsymbol{\nu}}(r(\boldsymbol{\theta})Z_0) \,|\, r(\boldsymbol{\theta}) \,] > 0 \quad \text{ on } \ \{r(\boldsymbol{\theta}) \neq 1\},$$

and thus

$$\mathbb{P}[\, k_{\boldsymbol{\nu}_0}(Z_0) \neq r(\boldsymbol{\theta}) k_{\boldsymbol{\nu}}(r(\boldsymbol{\theta})Z_0) \,] = \mathbb{E}[\, \mathbb{P}[k_{\boldsymbol{\nu}_0}(Z_0) \neq r(\boldsymbol{\theta}) k_{\boldsymbol{\nu}}(r(\boldsymbol{\theta})Z_0) \,|\, r(\boldsymbol{\theta})] \,] > 0,$$

which contradicts (6.9). Therefore necessarily $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and it remains to show that $k_{\boldsymbol{\nu}_0}(Z_0) = k_{\boldsymbol{\nu}}(Z_0)$ with probability 1 implies $\boldsymbol{\nu} = \boldsymbol{\nu}_0$. From $k_{\boldsymbol{\nu}_0}(Z_0) = k_{\boldsymbol{\nu}}(Z_0)$ a.s. and M.1 and M.2 we conclude $k_{\boldsymbol{\nu}_0}(x) = k_{\boldsymbol{\nu}}(x)$ for all $x \in \mathbb{R}$, so that by assumption M.3, $\boldsymbol{\nu} = \boldsymbol{\nu}_0$ follows. Altogether we have shown that $f_0(\boldsymbol{\tau}) = f_0(\boldsymbol{\tau}_0)$ a.s. if and only if $\boldsymbol{\tau} = \boldsymbol{\tau}_0$, or equivalently $L(\boldsymbol{\tau}) \leq L(\boldsymbol{\tau}_0)$ with equality if and only if $\boldsymbol{\tau} = \boldsymbol{\tau}_0$. This concludes the proof. □

Remark 6.1.3. One may replace the almost sure convergence in (6.7) by convergence in probability. Under the weaker assumption

$$n^{-1} \|\hat{L}_n - L_n\|_{K \times V} \xrightarrow{\mathbb{P}} 0, \qquad n \to \infty,$$

$\hat{\boldsymbol{\tau}}_n$ is *weakly* consistent, i.e., $\hat{\boldsymbol{\tau}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\tau}_0$. This follows because all the arguments used in Part 2 of the proof of Theorem 5.3.1 can be adapted if the pointwise almost sure convergence $n^{-1}\hat{L}_n(\boldsymbol{\theta}) \xrightarrow{\text{a.s.}} L(\boldsymbol{\theta})$ is replaced by $n^{-1}\hat{L}_n(\boldsymbol{\theta}) \xrightarrow{\mathbb{P}} L(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in K$. We omit details. □

A straightforward conclusion of Theorem 6.1.1 is the strong consistency of the MLE in GARCH(p,q). But first we appeal to condition Q.2 of Section 4.2.1 on the GARCH(p,q) or AGARCH(p,q) parameters, i.e., $\boldsymbol{\theta}_0 = (\alpha_0^\circ, \ldots, \alpha_p^\circ, \beta_1^\circ, \ldots, \beta_q^\circ)^T$ or $\boldsymbol{\theta}_0 = (\alpha_0^\circ, \ldots, \alpha_p^\circ, \beta_1^\circ, \ldots, \beta_q^\circ, \gamma^\circ)^T$, respectively:

Q.2 $(\alpha_p^\circ, \beta_q^\circ) \neq (0,0)$, $\alpha_0^\circ > 0$ and there is $i > 0$ with $\alpha_i^\circ > 0$. The polynomials $\alpha^\circ(z) = \alpha_1^\circ z + \cdots + \alpha_p^\circ z^p$ and $\beta^\circ(z) = 1 - \beta_1^\circ z - \cdots - \beta_q^\circ z^q$ do not have any common zeros.

We also remind the definition of the set B:

$$B = \left\{ (\beta_1, \ldots, \beta_q)^T \in [0,1)^q \,\middle|\, \sum_{j=1}^q \beta_j < 1 \right\}.$$

Since in GARCH(p,q) the condition Q.2 implies C.1 – C.4, the MLE in GARCH(p,q) is strongly consistent under M.1 – M.3.

Theorem 6.1.4. *Let (X_t) be a stationary* GARCH(p,q) *process with innovations obeying* (6.2) *and true parameter vector $\boldsymbol{\tau}_0 = \binom{\boldsymbol{\theta}_0}{\boldsymbol{\nu}_0}$. Take $K \subset (0,\infty) \times [0,\infty)^p \times B$ compact such that it contains $\boldsymbol{\theta}_0$ and assume that condition* Q.2 *applies to $\boldsymbol{\theta}_0$. Suppose the class of densities $\mathcal{D} = \{k_{\boldsymbol{\nu}} \,|\, \boldsymbol{\nu} \in V\}$ is such that* M.1 – M.3 *are fulfilled and properties* (6.6) – (6.7) *hold. Then the* MLE *$\hat{\boldsymbol{\tau}}_n$ is strongly consistent.*

In the proof of Theorem 5.4.6 we verified C.1 – C.4 for AGARCH(p,q). Therefore the following result is immediate.

Theorem 6.1.5. *Let (X_t) be a stationary* AGARCH(p,q) *process with innovations obeying* (6.2) *and true parameter vector $\boldsymbol{\tau}_0 = \binom{\boldsymbol{\theta}_0}{\boldsymbol{\nu}_0}$. Take $K \subset (0,\infty) \times [0,\infty)^p \times B \times [-1,1]$ compact such that it contains $\boldsymbol{\theta}_0$ and assume that condition* Q.2 *applies to $\boldsymbol{\theta}_0$. Suppose the class of densities $\mathcal{D} = \{k_{\boldsymbol{\nu}} \,|\, \boldsymbol{\nu} \in V\}$ is such that* M.1 – M.3 *are fulfilled and properties* (6.6) – (6.7) *hold. Then the* MLE *$\hat{\boldsymbol{\tau}}_n$ is strongly consistent.*

6.1.2 Consistency of the MLE with Respect to Student $\mathbf{t}_\nu$ Innovations

In this section we consider MLE with respect to Student t_ν innovations in model (6.1). The difficulty is to establish (6.7). The following criterion is tailored for the treatment of Student t_ν innovations.

Lemma 6.1.6. *Under* C.1 – C.4 *of Section 5.3, property* (6.7) *is fulfilled if there exist $c > 0$ and $0 < \kappa \leq 1$ such that for all $x, y \in \mathbb{R}$*

$$\sup_{\boldsymbol{\nu} \in K} |\log k_{\boldsymbol{\nu}}(x) - \log k_{\boldsymbol{\nu}}(y)| = \| \log k_{\bullet}(x) - \log k_{\bullet}(y) \|_V \leq c\,|x^2 - y^2|^\kappa. \tag{6.10}$$

Proof. Observe that

$$\frac{1}{n}\left(\hat{L}_n(\boldsymbol{\tau}) - L_n(\boldsymbol{\tau})\right) = \frac{1}{2n}\sum_{t=1}^n (\log h_t(\boldsymbol{\theta}) - \log \hat{h}_t(\boldsymbol{\theta}))$$
$$+ \frac{1}{n}\sum_{t=1}^n \Big(\log k_{\boldsymbol{\nu}}(\,(\hat{h}_t(\boldsymbol{\theta}))^{-1/2} X_t) - \log k_{\boldsymbol{\nu}}(\,(h_t(\boldsymbol{\theta}))^{-1/2} X_t)\Big)\,. \tag{6.11}$$

The proof of Theorem 5.3.1 shows that

$$\frac{1}{2n}\Big\| \sum_{t=1}^{n}(\log h_t - \log \hat{h}_t) \Big\|_K \xrightarrow{\text{a.s.}} 0, \qquad n \to \infty.$$

Consequently, we are left to treat the second sum (6.11), which we denote by $R_n(\boldsymbol{\tau})$. By virtue of inequality (6.10),

$$\begin{aligned}
\|R_n\|_{K\times V} &\le \frac{1}{n}\sum_{t=1}^{n} \|\log k_\bullet(\hat{h}_t^{-1/2} X_t) - \log k_\bullet(h_t^{-1/2} X_t)\|_{K\times V} \\
&\le \frac{c}{n}\sum_{t=1}^{n} X_t^2\, \|\hat{h}_t^{-1} - h_t^{-1}\|_K^{\kappa} \\
&\le \frac{\tilde{c}}{n}\sum_{t=1}^{n} X_t^2\, \|\hat{h}_t - h_t\|_K^{\kappa} \xrightarrow{\text{a.s.}} 0, \qquad n \to \infty.
\end{aligned}$$

The last inequality (with $\tilde{c} > 0$ some constant) is a consequence of the mean value theorem, and $\hat{h}_t, h_t \ge \underline{g} > 0$. The limit relation follows from the facts that $\|\hat{h}_t - h_t\|_K^{\kappa} \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$ and $\mathbb{E}(\log^+ X_0^2) < \infty$ together with an application of Proposition 2.5.1. This concludes the proof. □

Now we explicitly assume that the underlying distribution of Z_0 in model (6.1) is Student t_ν with unit variance, i.e., Z_0 has Lebesgue density

$$k_\nu(x) = c_1(\nu)\,(1 + c_2(\nu)\,x^2)^{-(\nu+1)/2}, \qquad x \in \mathbb{R}, \tag{6.12}$$

where

$$\left.\begin{aligned}
c_1(\nu) &= \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)}\sqrt{\frac{\nu}{\nu-2}}\,, \\
c_2(\nu) &= \frac{1}{\nu-2}\,.
\end{aligned}\right\} \quad (\nu > 2)$$

Observe that we have chosen $c_1(\nu)$ and $c_2(\nu)$ such that $\int_{-\infty}^{\infty} x^2 k_\nu(x)\,\mathrm{d}x = 1$. Let $V \subset (2,\infty)$ be compact and contain ν_0 in its interior. It is straightforward to see that $\mathcal{D} = \{k_\nu \mid \nu \in V\}$ fulfills M.1 – M.3 and assumption (6.6). Furthermore, since

$$\begin{aligned}
|\log k_\nu(x) - \log k_\nu(y)| &\le \frac{\nu+1}{2}\,\left|\log\left(1 + c_2(\nu)\,x^2\right) - \log\left(1 + c_2(\nu)\,y^2\right)\right| \\
&\le \frac{c_2(\nu)\,(\nu+1)}{2}\,|x^2 - y^2|\,,
\end{aligned}$$

the inequality (6.10) holds true with $c = \sup_{\nu\in V}\{(\nu+1)/(2\nu-4)\}$ and $\kappa = 1$. It is also easy to see that (6.6) holds true for a Student t_{ν_0} random variable.

Consequently, provided the conditions C.1 – C.4 of Section 5.3 are met, the MLE $\hat{\boldsymbol{\tau}}_n$ with respect to Student t_ν innovations is strongly consistent by virtue of Theorem 6.1.1. This insight leads to the consistency of the MLE in model (6.1) or (A)GARCH(p, q) with Student t_ν innovation.

Corollary 6.1.7. *Let* (X_t) *be a stationary process of form* (6.1) *with innovations obeying* (6.2), *where* $\mathcal{D}$ *consists of the Student* t_ν *densities* (6.12). *Denote by* $\boldsymbol{\tau}_0 = \binom{\boldsymbol{\theta}_0}{\nu_0}$ *and assume that the compact set* K *contains* $\boldsymbol{\theta}_0$ *and is such that conditions* C.1 – C.4 *of Section 5.3 hold. Suppose that the compact set* $V \subset (2, \infty)$ *contains* ν_0. *Then the* MLE *defined by* (6.4) *is strongly consistent, i.e.,* $\hat{\boldsymbol{\tau}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\tau}_0$.

Corollary 6.1.8. *Let* (X_t) *be a stationary* (A)GARCH(p, q) *process with innovations obeying* (6.2), *where* $\mathcal{D}$ *consists of the Student* t_ν *densities* (6.12). *Denote by* $\boldsymbol{\tau}_0 = \binom{\boldsymbol{\theta}_0}{\nu_0}$ *and assume that the compact set* $K \subset (0, \infty) \times [0, \infty)^p \times B$ *(or* $K \subset (0, \infty) \times [0, \infty)^p \times B \times [-1, 1]$, *respectively) contains* $\boldsymbol{\theta}_0$ *and that condition* Q.2 *applies to* $\boldsymbol{\theta}_0$. *Suppose that the compact set* V *contains* ν_0. *Then the* MLE $\hat{\boldsymbol{\tau}}_n$ *is strongly consistent, i.e.,* $\hat{\boldsymbol{\tau}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\tau}_0$.

6.2 Misspecification of the Innovations Density

Apart from artificially generated time series, it is impossible to specify the distribution of the innovations Z_t completely. In particular, we can never be certain that the distribution underlying Z_0 belongs to the class $\mathcal{D} = \{k_{\boldsymbol{\nu}} \mid \boldsymbol{\nu} \in V\}$. Assume that we have misspecified $\mathcal{D}$, i.e., there is no $\boldsymbol{\nu}_0 \in V$ such that Z_0 has Lebesgue density $k_{\boldsymbol{\nu}_0}$. In this situation it does not make sense to talk about consistency of $\hat{\boldsymbol{\tau}}_n = \binom{\hat{\boldsymbol{\theta}}_n}{\hat{\boldsymbol{\nu}}_n}$ because there is no true parameter $\boldsymbol{\nu}_0$. But as we are primarily interested in the parameter $\boldsymbol{\theta}$, we could still ask whether $\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$ as $n \to \infty$. Since the QMLE *is* consistent even though it is based on an incorrect likelihood (see Theorem 5.3.1), we might be tempted to conjecture that $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator for $\boldsymbol{\theta}$. However, this guess is wrong.

6.2.1 Inconsistency of the MLE

In the next theorem we study the MLE of $\boldsymbol{\theta}$ in a misspecified model.

Theorem 6.2.1. *Let* (X_t) *be a stationary process of form* (6.1) *with* (Z_t) *iid and* $\mathbb{E}Z_0 = 0$, $\mathbb{E}Z_0^2 = 1$. *Suppose the compact set* K *is such that the conditions* C.1 – C.4 *of Theorem 5.3.1 hold and that the true parameter* $\boldsymbol{\theta}_0$ *is contained in the interior of* K *and* $\mathbb{E}(h_0'(\boldsymbol{\theta}_0)/\sigma_0^2) \neq \mathbf{0}$. *Let the functions* $\hat{L}_n(\boldsymbol{\tau})$ *and* $L_n(\boldsymbol{\tau})$ *be as in* (6.3) *and* (6.5), *with respect to the class of Lebesgue densities*

$$\mathcal{D} = \{k_{\boldsymbol{\nu}} \mid \boldsymbol{\nu} \in V\}.$$

Define $\hat{\boldsymbol{\tau}}_n = \binom{\hat{\boldsymbol{\theta}}_n}{\hat{\boldsymbol{\nu}}_n}$ *as a maximizer of* $\hat{L}_n(\boldsymbol{\tau})$. *Suppose:*

1. *The class $\mathcal{D}$ obeys* M.1 – M.3.

2. *The function* $\mathbb{R} \times V \to (0, \infty)$: $(x, \boldsymbol{\nu}) \mapsto k_{\boldsymbol{\nu}}(x)$ *is continuously differentiable, for every* $\boldsymbol{\nu} \in V$ *the function* $x k'_{\boldsymbol{\nu}}(x)$ *(here* $k'_{\boldsymbol{\nu}}(x) = \partial k_{\boldsymbol{\nu}}(x)/\partial x$*) is integrable and*
$$k_{\boldsymbol{\nu}}(x) = o(|x|^{-1}), \qquad |x| \to \infty.$$

3. *The moment conditions*
$$\mathbb{E}\|\log f_0\|_{K\times V} < \infty \qquad \textit{and} \qquad \mathbb{E}\left\|\frac{\partial(\log f_0)}{\partial \boldsymbol{\tau}}\right\|_{K\times V} < \infty$$
hold and the following limit relations are valid:
$$n^{-1}\|\hat{L}_n - L_n\|_{K\times V} \xrightarrow{\text{a.s.}} 0 \qquad \textit{and} \qquad n^{-1}\left\|\frac{\partial \hat{L}_n}{\partial \boldsymbol{\tau}} - \frac{\partial L_n}{\partial \boldsymbol{\tau}}\right\|_{K\times V} \xrightarrow{\text{a.s.}} 0.$$

4. *For all* $\boldsymbol{\nu} \in V$*,*
$$\mathbb{E}\left(Z_0 \frac{k'_{\boldsymbol{\nu}}(Z_0)}{k_{\boldsymbol{\nu}}(Z_0)}\right) \neq -1. \tag{6.13}$$

Then the estimator $\hat{\boldsymbol{\theta}}_n$ *is inconsistent.*

Remark 6.2.2. Note that the condition $\mathbb{E}(h'_0(\boldsymbol{\theta}_0)/\sigma_0^2) \neq \mathbf{0}$ of the theorem is e.g. fulfilled in (A)GARCH(p, q) since there
$$\frac{\partial h_t(\boldsymbol{\theta})}{\partial \alpha_0} = \left(1 - \sum_{j=1}^{q} \beta_j\right)^{-1} > 0$$
for every $\boldsymbol{\theta}$. □

Proof. From integration by parts together with $k_{\boldsymbol{\nu}}(x) = o(|x|^{-1})$ as $|x| \to \infty$,
$$\int_{-\infty}^{\infty} x \frac{k'_{\boldsymbol{\nu}}(x)}{k_{\boldsymbol{\nu}}(x)} k_{\boldsymbol{\nu}}(x)\, \mathrm{d}x = \int_{-\infty}^{\infty} x\, k'_{\boldsymbol{\nu}}(x)\, \mathrm{d}x = -\int_{-\infty}^{\infty} k_{\boldsymbol{\nu}}(x)\, \mathrm{d}x = -1$$
for all $\boldsymbol{\nu} \in V$. Therefore Assumption 4 ensures that the class $\mathcal{D}$ is indeed misspecified. By virtue of Assumption 3 and an application of Theorem 2.2.1,
$$\|n^{-1}\hat{L}_n - L\|_{K\times V} \xrightarrow{\text{a.s.}} 0,$$
$$\|n^{-1}\hat{L}'_n - L'\|_{K\times V} \xrightarrow{\text{a.s.}} 0,$$
where

$$L(\boldsymbol{\tau}) = \mathbb{E}[\log f_0(\boldsymbol{\tau})],$$

$$L'(\boldsymbol{\tau}) = \frac{\partial L(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}} = \mathbb{E}\left(\frac{\partial \log f_0(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}}\right).$$

Assume by contradiction that $\hat{\boldsymbol{\theta}}_n$ is consistent, i.e., $\hat{\boldsymbol{\theta}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_0$ as $n \to \infty$. Since every random sequence converging in probability contains a subsequence converging almost surely, we may assume that $\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}_0$ for this proof. Choose a realization $\omega \in \Omega$ such that simultaneously $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ and $\|n^{-1}\hat{L}_n' - L'\|_{K\times V} \to 0$ as $n \to \infty$. Since V is compact, there is a convergent subsequence $\hat{\boldsymbol{\nu}}_{n_k}$ having a certain limit $\bar{\boldsymbol{\nu}} \in V$. Altogether we have the convergence $\hat{\boldsymbol{\tau}}_{n_k} \to \binom{\boldsymbol{\theta}_0}{\bar{\boldsymbol{\nu}}}$ as $k \to \infty$. Since $\boldsymbol{\theta}_0$ lies in the interior of K and $\hat{\boldsymbol{\nu}}_{n_k}$ maximizes L_{n_k}, we conclude that $\partial \hat{L}_n(\hat{\boldsymbol{\theta}}_{n_k})/\partial \boldsymbol{\theta} = \mathbf{0}$ for n_k large enough. Because of $\hat{\boldsymbol{\theta}}_{n_k} \to \boldsymbol{\theta}_0$ and $\|n_k^{-1}L_{n_k}' - L'\|_{K\times V} \to 0$,

$$\frac{\partial L}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0, \bar{\boldsymbol{\nu}}) \equiv \mathbf{0}.$$

On the other hand, observe that

$$\frac{\partial L(\boldsymbol{\tau})}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\mathbb{E}\left[\frac{h_0'(\boldsymbol{\theta})}{h_0(\boldsymbol{\theta})}\left(1 + \frac{k_{\boldsymbol{\nu}}'((h_0(\boldsymbol{\theta}))^{-1/2}X_0)\,(h_0(\boldsymbol{\theta}))^{-1/2}X_0}{k_{\boldsymbol{\nu}}((h_0(\boldsymbol{\theta}))^{-1/2}X_0)}\right)\right],$$

and hence (recall that $h_0(\boldsymbol{\theta}_0) = \sigma_0^2$ a.s. and that h_0, h_0' are independent of Z_0),

$$\mathbb{E}\left[\frac{h_0'(\boldsymbol{\theta}_0)}{\sigma_0^2}\left(1 + Z_0\frac{k_{\bar{\boldsymbol{\nu}}}'(Z_0)}{k_{\bar{\boldsymbol{\nu}}}(Z_0)}\right)\right] = \mathbb{E}\left(\frac{h_0'(\boldsymbol{\theta}_0)}{\sigma_0^2}\right)\mathbb{E}\left(1 + Z_0\frac{k_{\bar{\boldsymbol{\nu}}}'(Z_0)}{k_{\bar{\boldsymbol{\nu}}}(Z_0)}\right) = \mathbf{0}.$$

Since $\mathbb{E}(h_0'(\boldsymbol{\theta}_0)/\sigma_0^2) \neq \mathbf{0}$ by assumption, consistency of $\hat{\boldsymbol{\theta}}_n$ implies

$$\mathbb{E}\left(1 + Z_0\frac{k_{\bar{\boldsymbol{\nu}}}'(Z_0)}{k_{\bar{\boldsymbol{\nu}}}(Z_0)}\right) = 0.$$

This contradicts (6.13) and concludes the proof of the theorem. □

Theorem 6.2.1 is of little value as long as we do not have an answer to the question as to whether there are any distributions of Z_0 which obey (6.13). The problem is resolved by the following theorem: whenever $k_{\boldsymbol{\nu}}$ is not the standard normal density, it is possible to construct a random variable Z_0 with $\mathbb{E}Z_0 = 0$ and $\mathbb{E}Z_0^2 = 1$ which fulfills (6.13). At the same time the theorem may be seen as a characterization of the standard normal density.

Theorem 6.2.3. *Let $k(x)$ be a positive and twice differentiable Lebesgue density, which fulfills $\int_{-\infty}^{\infty} xk(x)\,\mathrm{d}x = 0$ and $\int_{-\infty}^{\infty} x^2k(x)\,\mathrm{d}x = 1$. Then*

$$\mathbb{E}\left(Z\frac{k'(Z)}{k(Z)}\right) = -1 \tag{6.14}$$

for every random variable Z with $\mathbb{E}Z = 0$ and $\mathbb{E}Z^2 = 1$ if and only if k is the standard normal density.

Proof. Clearly, $k(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ is sufficient for (6.14). For proving the necessity, we assume that $k(x) \neq \phi(x)$. The proof is completed if we can construct a random variable Z with $\mathbb{E}Z = 0$ and $\mathbb{E}Z^2 = 1$ such that $\mathbb{E}[q(Z)] \neq 0$, where

$$q(x) = x\left(\frac{k'(x)}{k(x)} + x\right).$$

There are three cases between which we distinguish.

Case 1: There is $a \geq 1$ such that $q(a) + q(-a) \neq 0$. Then for any random variable Z with distribution $\mathbb{P}(Z = a) = \mathbb{P}(Z = -a) = 1/(2a^2)$, $\mathbb{P}(Z = 0) = 1 - 1/a^2$, we have that $\mathbb{E}Z = 0$, $\mathbb{E}Z^2 = 1$, but $\mathbb{E}[q(Z)] = (q(a)+q(-a))/(2a^2) \neq 0$.

Case 2: For all $x \geq 1$: $q(x) + q(-x) = 0$ and $q(a) + q(-a) \neq 0$ for some $0 < a < 1$. We define a distribution by $\mathbb{P}(Z = a) = \mathbb{P}(Z = -a) = 1/4$ and $\mathbb{P}(Z = \sqrt{2 - a^2}) = \mathbb{P}(Z = -\sqrt{2 - a^2}) = 1/4$. Clearly, $\mathbb{E}Z = 0$, $\mathbb{E}Z^2 = 1$, but $\mathbb{E}[q(Z)] = (q(a) + q(-a))/4 \neq 0$, so that we are left with studying the last case.

Case 3: For all $x \in \mathbb{R}$: $q(x) + q(-x) = 0$, i.e., q is odd. We claim that there exists $a > 0$ with $q(a) \neq aq'(a)$. Indeed, $xq'(x) = q(x)$ for all $x \in \mathbb{R}$ would imply $q(x) = \lambda x$ for some $\lambda \in \mathbb{R}$, or in other words $k'(x)/k(x) = -x + \lambda$, from which we conclude $k(x) = C\exp(-x^2/2 + \lambda x)$ for some $C > 0$, i.e., k is a normal density. Since the mean of k is assumed to be zero, we end up with $\lambda = 0$ and $k(x) = \phi(x)$, which provides a contradiction to our assumption $k(x) \neq \phi(x)$. Thus $q(a) \neq aq'(a) \neq 0$ for at least one $a > 0$.

For the specification of a distribution of Z for which $\mathbb{E}[q(Z)] \neq 0$, we take $p = (2a^2 + 2)^{-1}$ and set $\epsilon = p\delta/a$, where $\delta \geq 0$ is a small number. Then we define $\mathbb{P}(Z = a + \delta) = p - \epsilon, \mathbb{P}(Z = -a + \delta) = p + \epsilon$, $\mathbb{P}(Z = b) = \mathbb{P}(Z = -b) = (1 - 2\epsilon)/2$, where $b > 0$ is chosen such that $\mathbb{E}Z^2 = 1$. We also note that $\mathbb{E}Z = 0$. The Taylor theorem together with the fact that q is odd yields

$$\begin{aligned}\mathbb{E}[q(Z)] &= (p - \epsilon)q(a + \delta) + (p + \epsilon)q(-a + \delta) \\ &= (p - \epsilon)q(a + \delta) - (p + \epsilon)q(a - \delta) \\ &= p(q(a + \delta) - q(a - \delta)) - \epsilon(q(a + \delta) + q(a - \delta)) \\ &= 2pq'(a)\delta - 2\epsilon q(a) + o(\delta) \\ &= 2p(q'(a) - q(a)/a)\delta + o(\delta), \qquad \delta \downarrow 0.\end{aligned}$$

Since $q'(a) \neq q(a)/a$, one can find a $\delta > 0$ such that $\mathbb{E}[q(Z)] \neq 0$. This completes the proof of the theorem. □

Remark 6.2.4. It is possible, but more involved, to construct a random variable Z having a Lebesgue density. The idea is to approximate the discrete distributions constructed above by piecewise constant densities. □

Remark 6.2.5. From a combination of Theorems 6.2.1 and 6.2.3, we recognize that in a model of form (6.1) where the innovations have true density $k_{\boldsymbol{\nu}_0}(x)$, even a tiny misspecification of the class $\mathcal{D} = \{k_{\boldsymbol{\nu}} \,|\, \boldsymbol{\nu} \in V\}$ of densities in general leads to inconsistent maximum likelihood estimates $\hat{\boldsymbol{\theta}}_n$. In contrast, the QMLE is always consistent (provided weak regularity assumptions hold, see Theorem 5.6.1). At this occasion, we clarify the meaning of a "tiny misspecification". Suppose that the family $\mathcal{D}$ fulfills regularity assumptions ensuring that one can construct a distribution function F such that for $Z_0 \sim F$,

$$\mathbb{E}\left(1 + Z_0 \frac{k'_{\boldsymbol{\nu}}(Z_0)}{k_{\boldsymbol{\nu}}(Z_0)}\right) \neq 0 \quad \text{for all } \boldsymbol{\nu} \in V.$$

This implies that in the so–called gross error model (cf. Huber [67])

$$Z_0 \sim (1-\epsilon) k_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x + \epsilon\, dF(x), \tag{6.15}$$

where $0 < \epsilon \leq 1$ is a contamination parameter and $\boldsymbol{\nu}_0 \in V$,

$$\mathbb{E}\left(1 + \frac{k'_{\boldsymbol{\nu}}(Z_0)}{k_{\boldsymbol{\nu}}(Z_0)}\right) \neq 0 \quad \text{for all } \boldsymbol{\nu} \in V.$$

Therefore, under the distributional assumption (6.15), the MLE based on $\mathcal{D}$ is inconsistent, even when $\epsilon > 0$ is very small.

Example 6.2.6. For pedagogical reasons we also illustrate the implications of Theorem 6.2.3 in a different, but closely related context. We study maximum likelihood estimation of scale in an iid sample. Consider the scale model

$$Y_t = \sigma Z_t, \tag{6.16}$$

where Z_t iid $\sim k_0(x)\,\mathrm{d}x$ and $\mathbb{E}Z_0 = 0$ and $\mathbb{E}Z_0^2 = 1$, i.e.,

$$\int_{-\infty}^{\infty} x k_0(x)\,\mathrm{d}x = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} x^2 k_0(x)\,\mathrm{d}x = 1.$$

$\sigma > 0$ is an unknown scale parameter. Suppose we observe $Y_1, \ldots, Y_n$ and misspecify the density k_0, i.e., we incorrectly base the estimation of σ on the log–likelihood with respect to a density $k_1 \neq k_0$, i.e., we maximize

$$L_n(\sigma) = -n \log(\sigma) + \sum_{t=1}^{n} \log k_1\left(Y_t/\sigma\right).$$

The maximizer of L_n can be seen as a quasi maximum likelihood estimator of scale.Also observe that k_0 and k_1 denote two *fixed* densities; there is no nuisance parameter $\boldsymbol{\nu}$.

Under regularity conditions on k_0 and k_1, uniformly on a compact set, $n^{-1}L_n \xrightarrow{\text{a.s.}} L$ and $n^{-1}L'_n \xrightarrow{\text{a.s.}} L'$ as $n \to \infty$, where

$$L(\sigma) = -\log(\sigma) + \mathbb{E}[\log k_1(Y_0/\sigma)],$$
$$L'(\sigma) = -\frac{1}{\sigma}\,\mathbb{E}\left(1 + \frac{Y_0}{\sigma}\,\frac{k_1'(Y_0/\sigma)}{k_1(Y_0/\sigma)}\right).$$

Denoting the true scale by σ° and following the arguments in the proof of Theorem 6.2.1, we recognize that $L'(\sigma^\circ) = 0$, i.e.,

$$\mathbb{E}\left(Z_0\,\frac{k_1'(Z_0)}{k_1(Z_0)}\right) = -1$$

is a *necessary* condition for the consistency of the QMLE of scale. As we have shown in Theorem 6.2.3, this condition is (in general) not fulfilled, unless $k_1(x) = \phi(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$. The choice $k_1 = \phi$ leads to the sample standard deviation $\hat{\sigma}_n$. In other words,

$$\hat{\sigma}_n = \left(\frac{1}{n}\sum_{t=1}^{n} Y_t^2\right)^{1/2} \tag{6.17}$$

is the only quasi maximum likelihood estimator of σ which is strongly consistent for *all* possible distributions of Z_0. For the scale model (6.16), the Gaussian QMLE (6.17) is in a sense "robust" with respect to misspecification of the density of Z_t, whereas the MLE is sensitive or "nonrobust". It is important to recognize that this notion does not coincide with classical definitions of robustness. The standard deviation (6.17) is e.g. not a qualitatively robust estimator, cf. Huber [67].

Related considerations were made by Gouriéroux et al. [60] in the framework of certain multivariate nonlinear regression models. They show that for such models quasi likelihood estimators are (strongly) consistent if and only if they are based on so–called quadratic exponential families. The scale model (6.16) also fits into their setup, and one might have the idea that their results are contradicted by our conclusions above. This discrepancy can however be resolved by observing that there is exactly one quadratic exponential family with respect to Lebesgue measure on $\mathbb{R}$: the Gaussian family.

There is another possible interpretation of Theorem 6.2.3. Since the work by Akaike [1] it has been well–known that maximum likelihood estimation of misspecified models (or quasi maximum likelihood estimation) asymptotically amounts to a minimization of the so–called Kullback–Leibler [79] discrepancy (KLD) between the true distribution and the model distributions. Under regularity assumptions (see e.g. White [132]), the QMLE converges a.s. to the model parameter which minimizes the KLD, and is in addition asymptotically normal. Possessing this information, the consequences of Theorem 6.2.3 may be alternatively formulated as follows:

> For data of form (6.16), the only scale model for which the KLD is *always* minimized at the true parameter σ°, is $X_t = \sigma Z_t$ with Z_t iid $\sim \mathcal{N}(0,1)$.

□

6.2.2 Misspecfication of $\mathcal{D}$ in the GARCH(p,q) Model

The problem of misspecfication of $\mathcal{D}$ in the GARCH(p,q) model was first described by Newey and Steigerwald [108] in the context of a general ARCH type model. To overcome the inconsistency problem, Newey and Steigerwald suggest to include an additional scale parameter σ, i.e., to compute the MLE with respect to the extended class

$$\tilde{\mathcal{D}} = \{\sigma^{-1} k_{\boldsymbol{\nu}}(\cdot/\sigma) \,|\, \boldsymbol{\nu} \in V, \sigma > 0\}.$$

They also bring up conditions under which consistency of the resulting estimator $\hat{\boldsymbol{\theta}}_n$ can be expected. Unfortunately these conditions are in general not satisfied by model (6.1)–(6.2) as their Assumption 2.4 is not even met by GARCH$(1,1)$. The reason is that the inclusion of an extra scale parameter leads to nonidentifiability.

Related problems concerning misspecification can be found in Berkes and Horváth [7]. They consider (non–Gaussian) QMLE with respect to a *fixed* density $k_0(x)$ in a GARCH(p,q) model. In their framework, no nuisance parameter $\boldsymbol{\nu}$ is jointly estimated with the GARCH parameters. Let us explain their main insight by means of the example of a Laplace density, standardized to unit variance, i.e.,

$$k_0(x) = 2^{-1/2}\, e^{-\sqrt{2}\,|x|}, \qquad x \in \mathbb{R}. \tag{6.18}$$

The quasi log–likelihood with respect to k_0 becomes

$$\hat{L}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} -\big(\log \hat{h}_t(\boldsymbol{\theta})^{1/2} + \log k_0(X_t/(\hat{h}_t(\boldsymbol{\theta}))^{1/2})\big).$$

Berkes and Horváth [7] demonstrate that the maximizer of the latter function is inconsistent iff $\mathbb{E}|Z_0| \neq 2^{-1/2}$. More precisely, $\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \bar{\boldsymbol{\theta}}_{0,d}$ as $n \to \infty$, where

$$\bar{\boldsymbol{\theta}}_{0,d} = \left(\frac{\alpha_0^\circ}{d^2}, \frac{\alpha_1^\circ}{d^2}, \ldots, \frac{\alpha_p^\circ}{d^2}, \beta_1^\circ, \ldots, \beta_q^\circ\right)^T$$

and $d = (\sqrt{2}\,\mathbb{E}|Z_0|)^{-1}$ (recall that $\boldsymbol{\theta}_0 = (\alpha_0^\circ, \ldots, \alpha_p^\circ, \beta_1^\circ, \ldots, \beta_q^\circ)^T$). Berkes and Horváth also establish asymptotic normality of $\hat{\boldsymbol{\theta}}_n$. A short argument for the special form of $\bar{\boldsymbol{\theta}}_{0,d}$ can be given as follows. By similar arguments as provided in the proof of Theorem 5.3.1, one shows that $\hat{L}_n/n \xrightarrow{\text{a.s.}} L$ in $\mathbb{C}(K)$, where the compact set K is as in Theorem 4.2.1 and

$$L(\boldsymbol{\theta}) = \mathbb{E}\left(\log\left[\frac{1}{\sqrt{h_0(\boldsymbol{\theta})}}\, k_0\left(\frac{\sqrt{h_0(\boldsymbol{\theta}_0)}Z_0}{\sqrt{h_0(\boldsymbol{\theta})}}\right)\right]\right).$$

Since the function $s \mapsto \mathbb{E}(\log[s k_0(sZ_0)])$ is uniquely maximized at $s = d$, the limit function L is maximal if and only if $h_0(\boldsymbol{\theta}_0)/h_0(\boldsymbol{\theta}) = d^2$ a.s. From representation (4.26) for $h_0(\boldsymbol{\theta})$ we deduce that the latter statement is equivalent to $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}_{0,d}$. Therefore $\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \bar{\boldsymbol{\theta}}_{0,d}$ as $n \to \infty$. If e.g. $Z_0 \sim \mathcal{N}(0,1)$, then $d = \sqrt{\pi}/2 \approx 0.89$. The following important example is in a similar spirit.

Quantification of the Error due to Misspecification of the Innovations Density

We consider the GARCH(p,q) model. The aim of this section is to determine the a.s. limit of the $\hat{\boldsymbol{\theta}}_n$ parameter in the MLE with respect to the Student t_ν density when there is misspecification. We appeal to the definition of the Student t_ν density in (6.12) and define

$$F(s,\nu) = \mathbb{E}[\log\{s k_\nu(sZ_0)\}] = \log s + \log c_1(\nu) - \frac{\nu+1}{2}\,\mathbb{E}[\log(1+c_2(\nu)s^2 Z_0^2)].$$

We need two auxiliary results.

Lemma 6.2.7. *There is a unique positive function $d(\nu)$ such that*

$$\sup_{s>0} F(s,\nu) = F(d(\nu),\nu) \quad \textit{for every } \nu > 2.$$

Proof. It is clear that $F(s,\nu) \to -\infty$ as $s \downarrow 0$. Moreover, since

$$\frac{\nu+1}{2}\,\mathbb{E}[\log(1+c_2(\nu)s^2Z_0^2)] \geq (\nu+1)\log s + \text{const}$$

for some constant not containing s, we have $F(s,\nu) \to -\infty$ as $s \to \infty$. Therefore the function $s \mapsto F(s,\nu)$ attains its maximum on $(0,\infty)$, say at $d(\nu)$. Since $\partial F(d(\nu),\nu)/\partial s = 0$, the point $d(\nu)$ obeys

$$(\nu+1)\,\mathbb{E}\left(\frac{c_2(\nu)\,d(\nu)Z_0^2}{1+c_2(\nu)\,d(\nu)^2 Z_0^2}\right) = \frac{1}{d(\nu)}$$

$$\Updownarrow$$

$$\mathbb{E}\left(\frac{c_2(\nu)\,d(\nu)^2 Z_0^2}{1+c_2(\nu)\,d(\nu)^2 Z_0^2}\right) = \frac{1}{\nu+1}$$

$$\Updownarrow$$

$$\mathbb{E}\left(\frac{1}{1+c_2(\nu)d(\nu)^2 Z_0^2}\right) = \frac{\nu}{\nu+1}\,.$$

From a monotonicity argument we then deduce that $d(\nu)$ is unique. □

We also employ the function $d(\nu)$ in the following lemma.

Lemma 6.2.8. *Let (X_t) be a stationary* GARCH(p,q) *process with parameter vector $\boldsymbol{\theta}_0 = (\alpha_0^\circ,\ldots,\alpha_p^\circ,\beta_1^\circ,\ldots,\beta_p^\circ)^T$. Take a compact set $K \subset (0,\infty)\times[0,\infty)^p \times B$ with $\boldsymbol{\theta}_0 \in K$ and assume that conditions* Q.1 *and* Q.2 *of Section 4.2.1 hold. Take $\nu > 2$ and suppose that*

$$\bar{\boldsymbol{\theta}}_{0,d(\nu)} = \left(\frac{\alpha_0^\circ}{d(\nu)^2}\,,\,\frac{\alpha_1^\circ}{d(\nu)^2}\,,\,\ldots\,,\,\frac{\alpha_p^\circ}{d(\nu)^2}\,,\,\beta_1^\circ,\ldots,\beta_q^\circ\right)^T \in K.$$

Then the function

$$\boldsymbol{\theta} \mapsto -\frac{1}{2}\mathbb{E}[\log h_0(\boldsymbol{\theta})] + \mathbb{E}[\log k_\nu((h_0(\boldsymbol{\theta}))^{-1/2}X_0)], \qquad \boldsymbol{\theta} \in K, \tag{6.19}$$

with ν fixed, is uniquely maximized at $\bar{\boldsymbol{\theta}}_{0,d(\nu)}$.

Proof. Adding the constant $2^{-1}\mathbb{E}[\log h_0(\boldsymbol{\theta}_0)]$ to the function (6.19), conditioning the expression on the right–hand side of (6.19) on h_0 and taking into account that h_0 is independent of Z_0, we note that maximization of (6.19) is equivalent to maximizing $\mathbb{E}[F(\{h_0(\boldsymbol{\theta}_0)/h_0(\boldsymbol{\theta})\}^{1/2}, \nu)]$ with respect to $\boldsymbol{\theta} \in K$. Lemma 6.2.7 implies

$$\mathbb{E}[F(\{h_0(\boldsymbol{\theta}_0)/h_0(\boldsymbol{\theta})\}^{1/2}, \nu)] \le F(d(\nu), \nu)$$

with equality if and only if

$$\frac{h_0(\boldsymbol{\theta}_0)}{h_0(\boldsymbol{\theta})} = d(\nu)^2 \quad \text{a.s.} \tag{6.20}$$

From representation (4.26) we conclude that (6.20) is equivalent to $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}_{0,d(\nu)}$. Indeed, sufficiency of $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}_{0,d(\nu)}$ for $h_0(\boldsymbol{\theta}_0)/h_0(\boldsymbol{\theta}) = d(\nu)^2$ is a trivial implication of the facts that the backshift operator $\beta_{\boldsymbol{\theta}}(B)$ is the same for both $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}_{0,d(\nu)}$ and that $\alpha_{\boldsymbol{\theta}}(B)$ is linear in $(\alpha_0, \ldots, \alpha_p)$; see Section 4.2.1 for the definitions of $\alpha_{\boldsymbol{\theta}}(z)$ and $\beta_{\boldsymbol{\theta}}(z)$. Necessity is a consequence of $h_0(\boldsymbol{\theta}) = h_0(\boldsymbol{\theta}_0)$ a.s. iff $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (Theorem 2.5 in Berkes et al. [8] or Lemma 5.4.5 with $\gamma^\circ = 0$). This completes the proof of the lemma. □

With evident notational modifications, the same statements are true in AGARCH(p, q) because the function $F(s, \nu)$ does not depend on the particular model and because $h_t(\boldsymbol{\theta})$ in AGARCH(p, q) is also linear in $(\alpha_1, \ldots, \alpha_p)$, see representation (5.43). In AGARCH(p, q), we have that $h_t(\boldsymbol{\theta}_0)/h_t(\boldsymbol{\theta}) = d(\nu)^2$ iff $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}_{0,d(\nu)} = (\alpha_0^\circ/d(\nu)^2, \ldots, \alpha_p^\circ/d(\nu)^2, \beta_1^\circ, \ldots, \beta_q^\circ, \gamma^\circ)^T$.

We now exploit Lemmas 6.2.7 and 6.2.8 in order to investigate the asymptotic behavior of the MLE $\hat{\boldsymbol{\theta}}_n$ in GARCH(p, q). We take the usual definition of the MLE, see (6.4) Also assume that the compact set K is given by (4.22) and Q.1 and Q.2 hold and that $V \subset (2, \infty)$ is compact. It is standard to show $n^{-1}\hat{L}_n \xrightarrow{\text{a.s.}} L = \mathbb{E}(\log f_0)$ in $\mathbb{C}(K \times V)$. Alternatively, we write $L(\boldsymbol{\theta}, \nu)$ for $L(\boldsymbol{\tau})$, $\boldsymbol{\tau} = \binom{\boldsymbol{\theta}}{\nu}$. The strategy we pursue is to first maximize L with respect to $\boldsymbol{\theta} \in K$, and in a second step with respect to $\nu \in V$. Lemma 6.2.8 tells us that for $\nu \in V$ fixed, $\boldsymbol{\theta} \mapsto L(\boldsymbol{\theta}, \nu)$ is uniquely maximized at $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}_{0,d(\nu)}$, in particular

$$\sup_{\boldsymbol{\theta} \in K} L(\boldsymbol{\theta}, \nu) = L(\bar{\boldsymbol{\theta}}_{0,d(\nu)}, \nu) = F(d(\nu), \nu) + \text{const},$$

where the constant does not depend on ν. Thus, if the function $\nu \mapsto F(d(\nu), \nu)$ is uniquely maximized at $\bar{\nu}_0 \in V$ and $\bar{\boldsymbol{\theta}}_{0,d(\bar{\nu}_0)} \in K$, then the function L is uniquely maximized at $L(\bar{\boldsymbol{\theta}}_{0,d(\bar{\nu}_0)}, \bar{\nu}_0)$, and by standard arguments,

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \bar{\boldsymbol{\theta}}_{0,d(\bar{\nu}_0)}, \qquad n \to \infty.$$

In exactly the same manner the AGARCH(p,q) case can be treated. Rather than formulating a mathematical theorem, we provide graphs of $F(d(\nu),\nu)$ and of $d(\nu)$ when $Z_0 \sim \mathcal{N}(0,1)$ in Figure 6.1 and when Z_0 has Laplace density (6.18) in Figure 6.2.

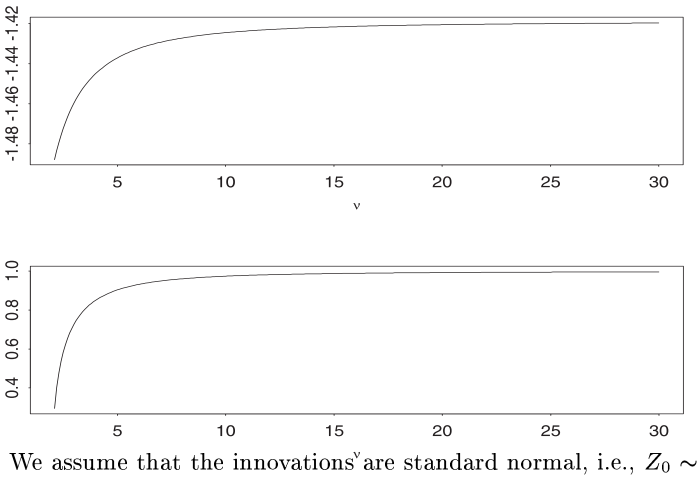

Fig. 6.1. We assume that the innovations are standard normal, i.e., $Z_0 \sim \mathcal{N}(0,1)$. *Top*: Graph of $F(d(\nu),\nu)$. *Bottom*: Graph of $d(\nu)$. The error is tiny, provided the set V contains large values, because $d(\nu) \to 1$ as $\nu \to \infty$. Observe that $(\hat{\nu}_n)$ converges to the right boundary of V as $n \to \infty$. This is not surprising in the light of the fact that the Student t_ν distribution tends to the standard normal distribution as $\nu \to \infty$.

6.3 Asymptotic Normality of the MLE

We now establish asymptotic normality of the MLE $\hat{\boldsymbol{\tau}}_n$ in model (6.1)–(6.2). The techniques are similar as in the proof of asymptotic normality of the QMLE, see Theorem 5.6.1. As the arguments heavily depend on a Taylor expansion of $L_n(\boldsymbol{\tau})$, we start by *formally* computing the derivatives of the function

$$\ell_0(\boldsymbol{\tau}) = \log f_0(\boldsymbol{\tau}) = \log s(\boldsymbol{\theta}) + \log k_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)\,,$$

where

$$s(\boldsymbol{\theta}) = (h_0(\boldsymbol{\theta}))^{-1/2}.$$

Since we primarily want to gain intuition, we are not yet concerned about technical details. Regularity assumptions will be imposed later, see below.

We suppose that (X_t) is the unique stationary ergodic solution to model (6.1)–(6.2) with $\boldsymbol{\tau} = \boldsymbol{\tau}_0 = \binom{\boldsymbol{\theta}_0}{\boldsymbol{\nu}_0}$. Assume that $h_0(\boldsymbol{\theta})$ and $k_{\boldsymbol{\nu}}(x)$ are twice continuously differentiable. We introduce the notation

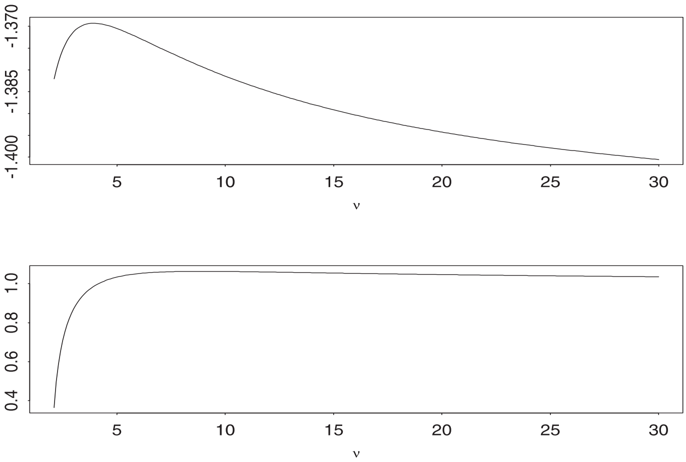

Fig. 6.2. We assume that the innovations Z_t have Laplace density (6.18). *Top*: Graph of $F(d(\nu),\nu)$. *Bottom*: Graph of $d(\nu)$. The maximum of $F(d(\nu),\nu)$ is attained at $\bar{\nu}_0 = 3.9$ (numerical approximation), and $d(\bar{\nu}_0) = 0.9841$. Therefore the error is tiny, provided $\bar{\nu}_0 \in V$. Moreover $\hat{\nu}_n \xrightarrow{\text{a.s.}} \bar{\nu}_0$, which may be regarded as rather peculiar since the Laplace distribution is light–tailed; we would also have expected that $(\hat{\nu}_n)$ converges to the right boundary of V.

$$\dot{k}_{\boldsymbol{\nu}}(x) = \frac{\partial k_{\boldsymbol{\nu}}(x)}{\partial \boldsymbol{\nu}} .$$

This function should not be confounded with $k'_{\boldsymbol{\nu}}(x)$, which is the first derivative of $k_{\boldsymbol{\nu}}(x)$ with respect to x. Correspondingly, $\dot{k}'_{\boldsymbol{\nu}}(x)$ is a mixed derivative.

We calculate that

$$\frac{\partial \ell_0}{\partial \boldsymbol{\theta}} = \frac{s'(\boldsymbol{\theta})}{s(\boldsymbol{\theta})} + \frac{k'_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)\, s'(\boldsymbol{\theta})X_0}{k_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)} , \tag{6.21}$$

$$\frac{\partial \ell_0}{\partial \boldsymbol{\nu}} = \frac{\dot{k}_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)}{k_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)} . \tag{6.22}$$

The matrices

$$\frac{\partial^2 \ell_0}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\nu}} = (s'(\boldsymbol{\theta}))^T \left[\frac{\partial}{\partial \boldsymbol{\nu}} \left(\frac{k'_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)}{k_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)} \right) \right] X_0 , \tag{6.23}$$

$$\frac{\partial^2 \ell_0}{\partial \boldsymbol{\nu}^2} = \frac{\partial}{\partial \boldsymbol{\nu}} \left(\frac{\dot{k}_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)}{k_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)} \right) \tag{6.24}$$

and

$$\frac{\partial^2 \ell_0}{\partial \boldsymbol{\theta}^2} = \frac{s''(\boldsymbol{\theta})s(\boldsymbol{\theta}) - (s'(\boldsymbol{\theta}))^T(s'(\boldsymbol{\theta}))}{(s(\boldsymbol{\theta}))^2}$$

$$+ \; \frac{k''_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)\,(s'(\boldsymbol{\theta}))^T s'(\boldsymbol{\theta})X_0^2 \; + \; k'_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)s''(\boldsymbol{\theta})X_0}{k_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)}$$

$$- \; \frac{(k'_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0))^2(s'(\boldsymbol{\theta}))^T s'(\boldsymbol{\theta})X_0^2}{(k_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0))^2}$$

$$= -\frac{(s'(\boldsymbol{\theta}))^T s'(\boldsymbol{\theta})}{(s(\boldsymbol{\theta}))^2}\left(1 + s(\boldsymbol{\theta})X_0\,\frac{k'_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)}{k_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)}\right)^2$$

$$+ \frac{(s'(\boldsymbol{\theta}))^T s'(\boldsymbol{\theta})}{(s(\boldsymbol{\theta}))^2}\left((s(\boldsymbol{\theta})X_0)^2\,\frac{k''_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)}{k_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)} + 2(s(\boldsymbol{\theta})X_0)\,\frac{k'_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)}{k_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)}\right)$$

$$+ \frac{s''(\boldsymbol{\theta})}{s(\boldsymbol{\theta})}\left(1 + (s(\boldsymbol{\theta})X_0)\,\frac{k'_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)}{k_{\boldsymbol{\nu}}(s(\boldsymbol{\theta})X_0)}\right) \tag{6.25}$$

form the entries of the Hessian of $\ell_0(\boldsymbol{\tau})$, i.e.,

$$\left[\frac{\partial^2 \ell_0}{\partial \boldsymbol{\tau}^2}\right] = \begin{bmatrix} \dfrac{\partial^2 \ell_0}{\partial \boldsymbol{\theta}^2} & \dfrac{\partial^2 \ell_0}{\partial \boldsymbol{\theta}\partial \boldsymbol{\nu}} \\[2ex] \dfrac{\partial^2 \ell_0}{\partial \boldsymbol{\nu}\partial \boldsymbol{\theta}} & \dfrac{\partial^2 \ell_0}{\partial \boldsymbol{\nu}^2} \end{bmatrix}.$$

We now introduce several matrices:

$$\mathbf{M}(\boldsymbol{\theta}) = \mathbb{E}\left(\frac{(s'(\boldsymbol{\theta}))^T s'(\boldsymbol{\theta})}{(s(\boldsymbol{\theta}))^2}\right) = \frac{1}{4}\,\mathbb{E}\left(\frac{(h'_0(\boldsymbol{\theta}))^T h'_0(\boldsymbol{\theta})}{(h_0(\boldsymbol{\theta}))^2}\right) \quad \in \mathbb{R}^{d\times d},$$

$$\mathbf{N}(\boldsymbol{\theta}) = \mathbb{E}\left(\frac{s'(\boldsymbol{\theta})}{s(\boldsymbol{\theta})}\right) = -\frac{1}{2}\,\mathbb{E}\left(\frac{h'_0(\boldsymbol{\theta})}{h_0(\boldsymbol{\theta})}\right) \quad \in \mathbb{R}^{1\times d}.$$

For $Z \sim k_{\boldsymbol{\nu}}(x)\,\mathrm{d}x$ we set

$$\mathbf{I}(\boldsymbol{\nu}) = \mathbb{E}\left[\left(\frac{\dot{k}_{\boldsymbol{\nu}}(Z)}{k_{\boldsymbol{\nu}}(Z)}\right)^T\left(\frac{\dot{k}_{\boldsymbol{\nu}}(Z)}{k_{\boldsymbol{\nu}}(Z)}\right)\right] \quad \in \mathbb{R}^{d'\times d'},$$

$$\mathbf{I}_s(\boldsymbol{\nu}) = \mathbb{E}\left[\left(1 + Z\,\frac{k'_{\boldsymbol{\nu}}(Z)}{k_{\boldsymbol{\nu}}(Z)}\right)^2\right] \quad \in \mathbb{R},$$

$$\mathbf{J}(\boldsymbol{\nu}) = \mathbb{E}\left[Z\,\frac{\partial}{\partial \boldsymbol{\nu}}\left(\frac{k'_{\boldsymbol{\nu}}(Z)}{k_{\boldsymbol{\nu}}(Z)}\right)\right] \quad \in \mathbb{R}^{1\times d'}.$$

Observe that $\mathbf{I}(\boldsymbol{\nu})$ is the Fisher information about $\boldsymbol{\nu}$ in the model of densities $\{k_{\boldsymbol{\nu}}(x) \,|\, \boldsymbol{\nu} \in V\}$. The matrix $I_s(\boldsymbol{\nu})$ is the Fisher information at $\sigma = 1$ in the scale model of densities $\{\sigma k_{\boldsymbol{\nu}}(\sigma x) \,|\, \sigma > 0\}$, where $\boldsymbol{\nu} \in V$ is *fixed.*

We now state regularity assumptions for model (6.1)–(6.2) and the true parameters $\boldsymbol{\theta}_0$, $\boldsymbol{\nu}_0$:

M.4 The assumptions N.1 and N.2 of Section 5.6 are fulfilled. Moreover, the matrices $\mathbf{M}(\boldsymbol{\theta}_0)$ and $\mathbf{N}(\boldsymbol{\theta}_0)$ are well–defined and $\mathbf{M}(\boldsymbol{\theta}_0)$ is positive definite.

M.5 The function $\mathbb{R} \times V \to (0, \infty)$: $(x, \boldsymbol{\nu}) \mapsto k_{\boldsymbol{\nu}}(x)$ is twice continuously differentiable in the interior of its domain.

M.6 The true parameter $\boldsymbol{\nu}_0$ lies in the interior of V and the density $k_{\boldsymbol{\nu}_0}$ is such that the following relations are true and the quantities appearing therein are well–defined for $Z_0 \sim k_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x$:

$$\mathbb{E}|\log k_{\boldsymbol{\nu}_0}(Z_0)| < \infty,$$

$$\mathbb{E}\left(\frac{\partial^2 \log k_{\boldsymbol{\nu}}(Z_0)}{\partial \boldsymbol{\nu}^2}(\boldsymbol{\nu}_0) \right) = -\mathbf{I}(\boldsymbol{\nu}_0), \tag{6.26}$$

$$\mathbb{E}\left(Z_0 \,\frac{k'_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)} \right) = -1, \tag{6.27}$$

$$\mathbb{E}\left(Z_0^2 \,\frac{k''_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)} \right) = -2\,\mathbb{E}\left(Z_0 \,\frac{k'_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)} \right) = 2, \tag{6.28}$$

$$\mathbb{E}\left(\frac{\dot{k}_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)} \right) = -\,\mathbb{E}\left(Z_0 \,\frac{\dot{k}'_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)} \right). \tag{6.29}$$

Moreover, $\mathbf{I}(\boldsymbol{\nu}_0)$ is positive definite.

M.7 For every $\mathbf{v} \in \mathbb{R}^{d'} \setminus \{\mathbf{0}\}$ the random variable

$$\left(1 + Z_0 \,\frac{k'_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)}\right)^{-1} \frac{\dot{k}_{\boldsymbol{\nu}_0}(Z_0)\,\mathbf{v}}{k_{\boldsymbol{\nu}_0}(Z_0)}$$

is nondegenerate.

M.8 The random element ℓ_0 obeys the following moment conditions:

$$\mathbb{E}\left\| \frac{\partial^2 \ell_0}{\partial \boldsymbol{\tau}^2} \right\|_{K \times V} < \infty, \qquad \mathbb{E}\left| \frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}} \right|^2 < \infty.$$

M.9 The following limit relations are satisfied:

$$n^{-1}\|\hat{L}_n - L_n\|_{K\times V} \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad n^{-1/2}\left\|\frac{\partial \hat{L}_n}{\partial \boldsymbol{\tau}} - \frac{\partial L_n}{\partial \boldsymbol{\tau}}\right\|_{K\times V} \xrightarrow{\text{a.s.}} 0.$$

Remark 6.3.1. Assumption M.4 guarantees that the function h_t is twice continuously differentiable and that $\boldsymbol{\theta}_0$ lies in the interior of K. Observe that the positive definiteness of $\mathbf{M}(\boldsymbol{\theta}_0)$ required in M.4 can be verified by checking Assumption N.4 of Section 5.6; see the proof of Lemma 5.6.3. Equation (6.26) in M.6 is a common regularity assumption about the Fisher information matrix, see e.g. Lehmann [85]. Relations (6.27)–(6.29) validate the following formal computations:

$$\begin{aligned}\mathbb{E}\left(Z_0\,\frac{k'_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)}\right) &= \int_{-\infty}^{\infty} x k'_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x \;=\; x k_{\boldsymbol{\nu}_0}(x)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} k_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x \\ &= -\int_{-\infty}^{\infty} k_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x = -1,\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left(Z_0^2\,\frac{k''_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)}\right) &= \int_{-\infty}^{\infty} x^2 k''_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x \;=\; x^2 k'_{\boldsymbol{\nu}_0}(x)\Big|_{-\infty}^{\infty} - 2\int_{-\infty}^{\infty} x\, k'_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x \\ &= -2\int_{-\infty}^{\infty} x k'_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x \\ &= -2\,\mathbb{E}\left(Z_0\,\frac{k'_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)}\right),\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left(\frac{\dot{k}_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)}\right) &= \int_{-\infty}^{\infty} \dot{k}_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x \;=\; x\dot{k}_{\boldsymbol{\nu}_0}(x)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} x\dot{k}'_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x \\ &= -\int_{-\infty}^{\infty} x\dot{k}'_{\boldsymbol{\nu}_0}(x)\,\mathrm{d}x \\ &= -\,\mathbb{E}\left(Z_0\,\frac{\dot{k}'_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)}\right).\end{aligned}$$

For verifying (6.27)–(6.29), one has to justify every step in the above calculations. M.7 is a technical assumption for which there does not seem to be any nice interpretation; M.7 will be exploited in the proof of the negative definiteness of $\mathbf{F}_0$ in Lemma 6.3.2 below. We will see in the proof of Theorem 6.3.3 below that the moment conditions M.8 are needed for the applications of Theorem 2.2.1 and the martingale central limit theorem. M.9 ensures that the maximizer of $L_n(\boldsymbol{\tau})$ is asymptotically equivalent to $\hat{\boldsymbol{\tau}}_n$. □

Lemma 6.3.2. *Under the conditions* M.1 – M.8, *the following matrix equalities are valid:*

$$\mathbf{F}_0 = \mathbb{E}\left[\frac{\partial^2 \ell_0}{\partial \boldsymbol{\tau}^2}(\boldsymbol{\tau}_0)\right] = -\,\mathbb{E}\left[\left(\frac{\partial \ell_0}{\partial \boldsymbol{\tau}}(\boldsymbol{\tau}_0)\right)^T\left(\frac{\partial \ell_0}{\partial \boldsymbol{\tau}}(\boldsymbol{\tau}_0)\right)\right]$$

$$= -\begin{bmatrix} \mathbf{I}_s(\boldsymbol{\nu}_0)\,\mathbf{M}(\boldsymbol{\theta}_0) & (\mathbf{N}(\boldsymbol{\theta}_0))^T\mathbf{J}(\boldsymbol{\nu}_0) \\ (\mathbf{J}(\boldsymbol{\nu}_0))^T\mathbf{N}(\boldsymbol{\theta}_0) & \mathbf{I}(\boldsymbol{\nu}_0) \end{bmatrix}. \tag{6.30}$$

Furthermore $\mathbf{F}_0$ *is negative definite.*

Proof. After noticing that

$$\left(\frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}}\right)^T \frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}} = \left[\frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\theta}}, \frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\nu}}\right]^T \left[\frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\theta}}, \frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\nu}}\right]$$

$$= \begin{bmatrix} \left(\frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\theta}}\right)^T \frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\theta}} & \left(\frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\theta}}\right)^T \frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\nu}} \\ \left(\frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\nu}}\right)^T \frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\theta}} & \left(\frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\nu}}\right)^T \frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\nu}} \end{bmatrix},$$

the identity (6.30) is established by routine calculations, where one takes into account that $s(\boldsymbol{\theta}_0) = 1/\sigma_0$, $s'(\boldsymbol{\theta}_0) = -2^{-1}h_0'(\boldsymbol{\theta}_0)/\sigma_0^3$ and $s''(\boldsymbol{\theta}_0)$ are independent of Z_0 and where one applies (6.26)–(6.29) to (6.21)–(6.25) with $\boldsymbol{\tau} = \boldsymbol{\tau}_0$.

It is evident that the matrix $\mathbf{F}_0$ is negative semi–definite. To show that it is also negative definite, one proceeds as in the proof of Lemma 5.6.3. Assume that there is $\mathbf{x} = \binom{\mathbf{u}}{\mathbf{v}}$ with $\mathbf{u} \in \mathbb{R}^d$ and $\mathbf{v} \in \mathbb{R}^{d'}$ such that $\mathbf{x}^T\mathbf{F}_0\mathbf{x} = \mathbf{0}$. This is equivalent to

$$\left(\frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}}\right)\mathbf{x} = \left(\frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\theta}}\right)\mathbf{u} + \left(\frac{\partial \ell_0(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\nu}}\right)\mathbf{v} = 0 \quad \text{a.s.}$$

or

$$-2^{-1}\left(1 + Z_0\,\frac{k'_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)}\right)\frac{h_0'(\boldsymbol{\theta}_0)\,\mathbf{u}}{\sigma_0^2} + \frac{\dot{k}_{\boldsymbol{\nu}_0}(Z_0)\,\mathbf{v}}{k_{\boldsymbol{\nu}_0}(Z_0)} = 0 \quad \text{a.s.} \tag{6.31}$$

The case $\mathbf{x} \neq \mathbf{0}$ leads to contradictions, as can be seen from an analysis of the following distinction of subcases:

$\{\mathbf{u} = \mathbf{0},\ \mathbf{v} \neq \mathbf{0}\}$: This contradicts the assumption M.6 that $\mathbf{I}(\boldsymbol{\nu}_0)$ is positive definite.

$\{\mathbf{u} \neq \mathbf{0},\ \mathbf{v} = \mathbf{0}\}$: By Assumption M.4, the matrix $\mathbf{M}(\boldsymbol{\theta}_0)$ is positive definite, and hence $\mathbb{P}(h_0'(\boldsymbol{\theta}_0)\mathbf{u}/\sigma_0^2 \neq \mathbf{0}) > 0$. It is also evident that $\mathbb{P}(1 + Z_0 k'_{\boldsymbol{\nu}_0}(Z_0)/k_{\boldsymbol{\nu}_0}(Z_0) = 0) < 1$. But this contradicts the fact that the random variables $1 + Z_0 k'_{\boldsymbol{\nu}_0}(Z_0)/k_{\boldsymbol{\nu}_0}(Z_0)$ and $h_0'(\boldsymbol{\theta}_0)\mathbf{u}/\sigma_0^2$ are independent.

$\{\mathbf{u} \neq \mathbf{0},\ \mathbf{v} \neq \mathbf{0}\}$: This would mean that

$$\frac{h_0'(\boldsymbol{\theta}_0)\,\mathbf{u}}{\sigma_0^2} = 2\left(1 + Z_0\,\frac{k'_{\boldsymbol{\nu}_0}(Z_0)}{k_{\boldsymbol{\nu}_0}(Z_0)}\right)^{-1} \frac{\dot{k}_{\boldsymbol{\nu}_0}(Z_0)\,\mathbf{v}}{k_{\boldsymbol{\nu}_0}(Z_0)} \quad \text{a.s.}$$

The right–hand side of the latter equation is nondegenerate by Assumption M.7. This however contradicts the independence between Z_0 and $h_0'(\boldsymbol{\theta}_0)\mathbf{u}/\sigma_0^2$ (Lemma 5.4.2).

Hence we have shown that (6.31) implies $\mathbf{u} = \mathbf{v} = \mathbf{0}$, which proves that $\mathbf{F}_0$ is indeed negative definite. This completes the proof. □

We are now ready to tackle asymptotic normality.

Theorem 6.3.3. *Let* (X_t) *be a stationary process of form* (6.1)–(6.2) *and true parameter vector* $\boldsymbol{\tau}_0$. *Under the assumptions* M.1 – M.9 *the* MLE $\hat{\boldsymbol{\tau}}_n$ *is asymptotically normal, i.e.,*

$$\sqrt{n}\,(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0) \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0}, -\mathbf{F}_0^{-1}), \qquad n \to \infty, \tag{6.32}$$

where $\mathbf{F}_0$ *is defined by* (6.30).

Proof. Observe that M.1 – M.9 imply that one can apply Theorem 6.1.1 so that $\hat{\boldsymbol{\tau}}_n \xrightarrow{\mathrm{a.s.}} \boldsymbol{\tau}_0$. From M.9 we can deduce analogously to the proof of Lemma 5.6.5 that the maximizer $\tilde{\boldsymbol{\tau}}_n$ of L_n is asymptotically equivalent to $\hat{\boldsymbol{\tau}}_n$, in particular even $\sqrt{n}\,|\hat{\boldsymbol{\tau}}_n - \tilde{\boldsymbol{\tau}}_n| \xrightarrow{\mathrm{a.s.}} 0$. Consequently it is enough to prove (6.32) with $\hat{\boldsymbol{\tau}}_n$ replaced by $\tilde{\boldsymbol{\tau}}_n$. By arguments identical to the proof of Theorem 5.6.1,

$$n^{1/2}(\tilde{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0) = -\mathbf{F}_0^{-1}\,(1 + o_{\mathbb{P}}(1))\,n^{-1/2}\,\frac{\partial L_n(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}}, \qquad n \to \infty, \tag{6.33}$$

By an application of the central limit theorem for finite variance stationary ergodic martingale difference sequences (Theorem 18.3 in Billingsley [12]) together with relation (6.30),

$$n^{-1/2}\,\frac{\partial L_n(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}} = n^{-1/2}\sum_{t=1}^{n}\frac{\partial \ell_t(\boldsymbol{\tau}_0)}{\partial \boldsymbol{\tau}} \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0},\ -\mathbf{F}_0), \qquad n \to \infty. \tag{6.34}$$

A combination of (6.33) and (6.34) followed by an application of Slutsky's lemma gives

$$\sqrt{n}\,(\tilde{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0) \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0}, -\mathbf{F}_0^{-1}).$$

This completes the proof. □

Remarks 6.3.4. The inverse of a partitioned matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix},$$

is equal to

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{\Sigma}^{11} & \mathbf{\Sigma}^{12} \\ \mathbf{\Sigma}^{21} & \mathbf{\Sigma}^{22} \end{pmatrix}$$

where

$$\begin{aligned} \mathbf{\Sigma}^{11} &= (\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})^{-1}, \\ \mathbf{\Sigma}^{12} &= -\mathbf{\Sigma}^{11}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}, \\ \mathbf{\Sigma}^{21} &= -\mathbf{\Sigma}^{22}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}, \\ \mathbf{\Sigma}^{22} &= (\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}; \end{aligned}$$

see e.g. Magnus and Neudecker [91]. An application of these formulas to $-\mathbf{F}_0$ shows that the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_n$ equals

$$\mathbf{\Sigma}_{\boldsymbol{\theta}_0}^{-1} = [\mathbf{I}_s(\boldsymbol{\nu}_0)\mathbf{M}(\boldsymbol{\theta}_0) - (\mathbf{N}(\boldsymbol{\theta}_0))^T\mathbf{J}(\boldsymbol{\nu}_0)(\mathbf{I}(\boldsymbol{\nu}_0))^{-1}(\mathbf{J}(\boldsymbol{\nu}_0))^T\mathbf{N}(\boldsymbol{\theta}_0)]^{-1}. \quad (6.35)$$

Since $[\mathbf{I}(\boldsymbol{\nu}_0)]^{-1}$ is positive definite by assumption M.6, also the matrix $(\mathbf{N}(\boldsymbol{\theta}_0))^T\mathbf{J}(\boldsymbol{\nu}_0)(\mathbf{I}(\boldsymbol{\nu}_0))^{-1}(\mathbf{J}(\boldsymbol{\nu}_0))^T$ is positive semi–definite and thus

$$\mathbf{I}_s(\boldsymbol{\nu}_0)\mathbf{M}(\boldsymbol{\theta}_0) \geq \mathbf{\Sigma}_{\boldsymbol{\theta}_0}, \quad (6.36)$$

where the notation $\mathbf{A} \geq \mathbf{B}$ for two square matrices means that $\mathbf{A} - \mathbf{B}$ is positive semi–definite. Observe also that one has equality in (6.36) if and only if $(\mathbf{N}(\boldsymbol{\theta}_0))^T\mathbf{J}(\boldsymbol{\nu}_0) = \mathbf{0}$, By Theorem 24 in Chapter 1 of Magnus and Neudecker [91], relation (6.36) is equivalent to

$$[\mathbf{I}_s(\boldsymbol{\nu}_0)\mathbf{M}(\boldsymbol{\theta}_0)]^{-1} \leq \mathbf{\Sigma}_{\boldsymbol{\theta}_0}^{-1}, \quad (6.37)$$

and one has strict inequality if and only if $(\mathbf{N}(\boldsymbol{\theta}_0))^T\mathbf{J}(\boldsymbol{\nu}_0) = \mathbf{0}$. Also observe that the matrix $[\mathbf{I}_s(\boldsymbol{\nu}_0)\mathbf{M}(\boldsymbol{\theta}_0)]^{-1}$ can be interpreted as the asymptotic covariance matrix of the MLE $\hat{\boldsymbol{\theta}}_n$ based on the *true* density $k_{\boldsymbol{\nu}_0}$. Therefore (6.37) implies that the inclusion of an unknown nuisance parameter $\boldsymbol{\nu}$ in general increases the asymptotic covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\theta}_0}^{-1}$ of $\hat{\boldsymbol{\theta}}_n$ as compared to MLE based on fixed and known $k_{\boldsymbol{\nu}_0}$. The situation where $\mathbf{\Sigma}_{\boldsymbol{\theta}_0}^{-1} > [\mathbf{I}_s(\boldsymbol{\nu}_0)\mathbf{M}(\boldsymbol{\theta}_0)]^{-1}$ is referred to as the estimator $\hat{\boldsymbol{\theta}}_n$ being *non*–adaptive for $\boldsymbol{\theta}$, see e.g. Bickel et al. [10]. In the context of adaptiveness in GARCH, one has to mention the articles by Drost and Klaassen [42] and Ling and McAleer [88]. These authors show that the GARCH(p,q) parameters $\boldsymbol{\theta}$ cannot be adaptively estimated and that one can construct adaptive estimators for $(\alpha_1, \alpha_2, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^T$.

We mention that formula (6.35) can be used for determining the asymptotic relative efficiency of the QMLE with respect to the MLE. It seems that in general one has to rely on simulation methods for determining $\mathbf{F}_0$ or $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}$. The asymptotic covariance matrix $-\mathbf{F}_0^{-1}$ (or $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}^{-1}$) can be consistently estimated from data, cf. Remark 5.6.2. We omit details because the form of this estimator is obvious. □

6.4 Asymptotic Normality of the MLE with Respect to Student t_ν Innovations

We continue the discussion of Section 6.1.2. Our task is to verify M.1 – M.9 in model (6.1) with iid Student t_ν innovations. Recall that

$$k_\nu(x) = c_1(\nu)\,(1 + c_2(\nu)\,x^2)^{-(\nu+1)/2}, \qquad x \in \mathbb{R},$$

where for $\nu > 2$,

$$c_1(\nu) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)}\sqrt{\frac{\nu}{\nu-2}},$$
$$c_2(\nu) = \frac{1}{\nu-2}.$$

For the sake of clarity, we devise assumptions which will be sufficient for M.1 – M.9. For the unique stationary solution (X_t) to model (6.1)–(6.2) with iid Student t_{ν_0} innovations and $\mathcal{D} = \{k_\nu \mid \nu \in V\}$ we assume:

T.1 Conditions N.1, N.2, N.4 of Section 5.6 hold.

T.2 The compact set $V \subset (2, \infty)$ contains the true parameter ν_0 in its interior.

T.3 The following moment conditions are met:

(i) $\mathbb{E}\left\| \frac{h_0'}{h_0} \right\|_K^2 < \infty$,

(ii) $\mathbb{E}\left\| \frac{h_0''}{h_0} \right\|_K < \infty$,

(iii) $\mathbb{E}\left\| \frac{\sigma_0^2 |h_0'|^2}{h_0^3} \right\|_K < \infty$,

(iv) $\mathbb{E}\left\| \frac{\sigma_0 |h_0'|^3}{h_0^{7/2}} \right\|_K < \infty$,

(v) $\mathbb{E}\left\| \frac{\sigma_0 |h_0'|^2}{h_0^{5/2}} \right\|_K < \infty$,

(vi) $\mathbb{E}\left\| \frac{\sigma_0 h_0''}{h_0^{3/2}} \right\|_K < \infty$.

We only sketch how M.1 – M.9 can be verified since the complete arguments are lengthy and not very instructive. The validity of M.1 – M.3 is evident. As regards M.4, the moment assumptions in T.3 imply that $\mathbf{M}(\boldsymbol{\theta}_0)$ and $\mathbf{N}(\boldsymbol{\theta}_0)$ are well–defined. Since T.1 contains N.4 of Section 5.6, by the arguments in the proof of Lemma 5.6.3 the matrix $\mathbf{M}(\boldsymbol{\theta}_0)$ is positive definite. M.6 and M.7 follow from straightforward arguments. Concerning M.8, one recognizes that there are constants $C_1, C_2 > 0$ such that for all $x \in \mathbb{R}$ and $\nu \in V$,

$$\left|\frac{\partial \log k_\nu(x)}{\partial \nu}\right| \le C_1 \log(1 + C_2 x^2), \qquad \left|\frac{\partial \log k_\nu(x)}{\partial x}\right| \le C_1,$$

$$\left|\frac{k''_\nu(x)}{k_\nu(x)}\right| \le C_1, \qquad \left|\frac{\partial^2 \log k_\nu(x)}{\partial x \partial \nu}\right| \le C_1,$$

$$\left|\frac{\partial^2 \log k_\nu(x)}{\partial \nu^2}\right| \le C_1, \qquad \left|\frac{\partial^2 \log k_\nu(x)}{\partial x^2}\right| \le C_1.$$

By using these upper bounds together with the moment conditions in T.3 one establishes M.8. The limit relations in M.9 are shown by means of application of the mean value theorem to $\hat{\ell}_t(\boldsymbol{\tau}) - \ell_t(\boldsymbol{\tau})$ and $(\partial/\partial\boldsymbol{\tau})(\hat{\ell}_t(\boldsymbol{\tau}) - \ell_t(\boldsymbol{\tau}))$, respectively, together with similar arguments as in the proof of Theorem 5.3.1 and Lemma 5.6.4. Thus we obtain the following corollary of Theorem 6.3.3.

Corollary 6.4.1. *Let* (X_t) *be a stationary process of form* (6.1)–(6.2) *with true parameter vector* $\boldsymbol{\tau}_0 = \binom{\boldsymbol{\theta}_0}{\boldsymbol{\nu}_0}$. *Let the class of densities* $\mathcal{D}$ *and the compact sets* K *and* V *be such that the conditions* T.1 – T.3 *are met. Then the* MLE $\hat{\boldsymbol{\tau}}_n$ *is strongly consistent and*

$$\sqrt{n}\,(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0) \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0}, -\mathbf{F}_0^{-1}), \qquad n \to \infty,$$

where the matrix $\mathbf{F}_0$ *is as in Lemma 6.3.2.*

The case of (A)GARCH(p,q) with Student t_ν innovations is quite relevant for applications. For GARCH (or AGARCH) we replace T.1 by Q.2 and Q.3 of Section 4.2.1. Observe that Q.1 and Q.4 are automatically fulfilled if Z_0 has a Student t_{ν_0} distribution. Q.1 – Q.4 in turn imply N.1, N.2 and N.4, i.e. condition T.1. For the verification of T.3 we recall that in (A)GARCH(p,q) with a Student t_{ν_0} distribution $\|h'_0/h_0\|_K$ and $\|h''_0/h_0\|_K$ possess finite moments of *any* order (Lemma 5.7.4) Moreover, $\mathbb{E}\|\sigma_0^2/h_0\|_K^s < \infty$ provided $s \in [0, \nu/2)$ by Lemma 5.7.5. With these two moment properties and Hölder's inequality one establishes the finite moments required in T.3. Consequently the following corollary holds true.

Corollary 6.4.2. *Let* (X_t) *be a stationary* (A)GARCH(p,q) *process with* Z_t *iid* $\sim t_{\nu_0}$. *Let the compact set* K *be as in Theorem 4.2.1 and assume the conditions* Q.2 *and* Q.3 *of Section 4.2.1 are met. Take* $V \subset (2, \infty)$ *compact such* *that* $\nu_0 \in V$. *Then the* MLE $\hat{\boldsymbol{\tau}}_n$ *with respect to the* t_ν *density is asymptotically normal, i.e.,*

$$\sqrt{n}\,(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0) \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0}, -\mathbf{F}_0^{-1}), \qquad n \to \infty,$$

where the matrix $\mathbf{F}_0$ *is as in Lemma 6.3.2.*

For illustration we attach Table 6.1, which contains asymptotic relative efficiencies of the QMLE with respect to the MLE of $\boldsymbol{\theta}$ in AGARCH(1, 1) models with t_ν innovations. We provide the asymptotic relative efficiencies for each parameter since in general there is no $\lambda > 0$ with $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}^{-1} = \lambda \mathbf{M}(\boldsymbol{\theta}_0)^{-1}$ (see formula (6.35)). In Figure 6.3 we compare the finite sample distributions of the QMLE and MLE for β_1.

ν	α_0	α_1	β_1	γ_1
5	0.277	0.105	0.247	0.509
9	0.786	0.834	0.805	0.829

Table 6.1. Asymptotic relative efficiencies of the QMLE with respect to the MLE in an AGARCH(1, 1) model with Student t_5 innovations and parameters $\alpha_0^\circ = 1.492 \times 10^{-6}$, $\alpha_1^\circ = 0.06706$, $\beta_1^\circ = 0.9079$ and $\gamma_1^\circ = 0.3594$ (the same parameters as used in Section 5.10). The numbers are ratios of means of estimated standard errors in 500 independent replications of the AGARCH(1, 1) time series with 10000 observations.

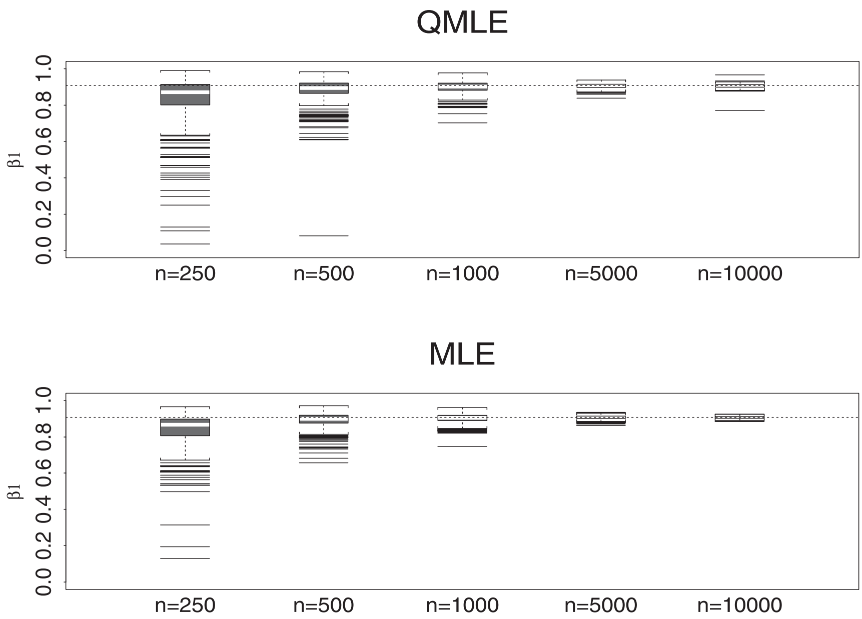

Fig. 6.3. AGARCH(1,1) model with Student t_5 innovations and parameters $\alpha_0^\circ = 1.492 \times 10^{-6}$, $\alpha_1^\circ = 0.06706$, $\beta_1^\circ = 0.9079$ and $\gamma_1^\circ = 0.3594$. For various sample sizes n, boxplots of 500 independent realizations of the QMLE and MLE of β_1 are drawn. The graphs show that it is beneficial to apply the MLE for any of the considered sample sizes n.

7

Quasi Maximum Likelihood Estimation in a Generalized Conditionally Heteroscedastic Time Series Model with Heavy–tailed Innovations

In this chapter we study the asymptotic behavior of the QMLE in the general heteroscedastic time series model (5.1) when the innovations are heavy in the sense that $\mathbb{E}Z_0^4 = \infty$. It turned out in the course of the proof of asymptotic normality (Theorem 5.6.1) that the asymptotic behavior of the QMLE $\hat{\boldsymbol{\theta}}_n$ is essentially determined by the limiting behavior of

$$L_n'(\boldsymbol{\theta}_0) = \frac{1}{2} \sum_{t=1}^{n} \frac{h_t'(\boldsymbol{\theta}_0)}{\sigma_t^2} (Z_t^2 - 1).$$

In this context,

$$\mathbf{H}_t = \frac{h_t'(\boldsymbol{\theta}_0)}{\sigma_t^2}, \qquad t \in \mathbb{Z},$$

is a stationary ergodic sequence of random vectors, which is adapted to the filtration

$$\mathcal{H}_t = \sigma(Y_s,\ s < t),$$

where $Y_t = Z_t^2 - 1$ constitutes an iid sequence with mean zero. If $\mathbf{H}_0$ has a finite first moment, the sequence $(\mathbf{H}_t Y_t)$ is a transform of the martingale difference sequence (Y_t), hence a stationary ergodic martingale difference sequence with respect to $(\mathcal{H}_t)$. If $\mathbb{E}|\mathbf{H}_0|^2 < \infty$ and $\mathbb{E}Y_0^2 < \infty$, an application of the central limit theorem for finite variance stationary ergodic martingale differences (see Billingsley [13], Theorem 18.3) yields

$$n^{-1/2} \sum_{t=1}^{n} \mathbf{H}_t Y_t \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{7.1}$$

where $\boldsymbol{\Sigma} = \mathbb{E}(Y_0^2\, \mathbf{H}_0^T \mathbf{H}_0)$ is the covariance matrix of $\mathbf{H}_0 Y_0$. This result does not require any additional information about the dependence structure of $(\mathbf{H}_t Y_t)$. For the asymptotic normality of the QMLE shown in Section 5.6 we imposed among others the moment conditions N.3 (i,ii), i.e., $\mathbb{E}Z_0^4 < \infty$ and $\mathbb{E}|\mathbf{H}_0|^2 < \infty$, which is equivalent to $\mathbb{E}(Y_0^2\, |\mathbf{H}_0|^2) < \infty$. It is needless to say that the purpose of N.3 (i,ii) was the applicability of the martingale central limit theorem (7.1).

If $\mathbb{E}(Y_0^2\,|\mathbf{H}_0|^2) = \infty$ a result as general as the CLT for stationary ergodic martingale differences is not known. However, some limit results for stationary sequences with marginal distributions in the domain of attraction of an infinite variance stable distribution exist. Recently, Hall and Yao [63] gave the asymptotic theory for QMLE in GARCH models when $\mathbb{E}Z_0^4 = \infty$. To be more specific, they assume regular variation with index $\kappa \in (1,2)$ for the distribution of Z_0^2. The aim of this chapter is to reprove these results in the setting of the general conditionally heteroscedastic time series model (5.1). The theory is based on a general limit result for the martingale transforms $\sum_{t=1}^n \mathbf{H}_t\, Y_t$ when the iid noise (Y_t) is regularly varying with index $\kappa \in (1,2)$, which was obtained by Mikosch and Straumann [103]. In contrast to Hall and Yao [63], the asymptotic theory for the QMLE presented here is not restricted to GARCH processes. The main difficulty of our approach is the verification of certain mixing conditions. In contrast to the case of asymptotic normality, such conditions cannot be avoided with our approach. It is difficult to check for a given model that these conditions hold; see Section 7.4 in order to get a flavor of the task to be solved. This chapter is based on Mikosch and Straumann [103].

7.1 Stable Limits of Infinite Variance Martingale Transforms

The literature on central limit theorems for martingales with infinite variance seems to be sparse. To the best of our knowledge, Mikosch and Straumann [103] are the only authors who provide some general results for a special class of martingales, so–called martingale transforms of a random walk. Assume $(\mathbf{Y}_t)$ has the particular form

$$\mathbf{Y}_t = \mathbf{H}_t Y_t, \qquad t \in \mathbb{Z},$$

where (Y_t) is an iid sequence and $(\mathbf{H}_t)$ is a strictly stationary sequence of random vectors with values in $\mathbb{R}^d$ such that $(\mathbf{H}_t)$ is adapted to the filtration given by the σ–fields $\mathcal{H}_t = \sigma(Y_s,\, s < t)$. If $\mathbb{E}Y_0 = 0$ and $\mathbb{E}|\mathbf{H}_0| < \infty$, then $\mathbb{E}(\mathbf{H}_t Y_t \,|\, \mathcal{H}_{t-1}) = \mathbf{0}$ a.s., and therefore $(\mathbf{H}_t)$ is a martingale difference sequence. The sequence

$$\mathbf{S}_0 = \mathbf{0}, \qquad \mathbf{S}_n = \mathbf{Y}_1 + \cdots + \mathbf{Y}_n, \quad n \geq 1,$$

is the martingale transform of the martingale $(\sum_{t=1}^n Y_t)_{n\geq 1}$ by the sequence $(\mathbf{H}_t)$. We keep this name even if $\mathbb{E}|\mathbf{Y}_0| = \infty$. We now introduce several assumptions:

A.1 Y_0 is regularly varying with index $\kappa \in (0,2)$.

A.2 There is some $\delta > 0$ with $\mathbb{E}|\mathbf{H}_0|^{\kappa+\delta} < \infty$.

A.3 The sequence $(\mathbf{Y}_t) = (\mathbf{H}_t Y_t)$ is strongly mixing with a geometric rate.

A random variable X is called regularly varying with index $\kappa \geq 0$ if there exist a probability $p \in [0,1]$ and a slowly varying function L such that indexgslowly varying

$$\mathbb{P}(X > x) = p\,\frac{L(x)}{x^\kappa} \quad \text{and} \quad \mathbb{P}(X < -x) = (1-p)\,\frac{L(x)}{x^\kappa}\,, \qquad x \to \infty.$$

L slowly varying means that $\lim_{x\to\infty} L(tx)/L(x) = 1$ for all $t > 0$. As an example, Pareto distributed random variables are regularly varying with $L(x) = \text{const.}$, cf. Embrechts et al. [45]. Another example is provided by the stationary AGARCH$(1,1)$ process (X_t), where the unconditional marginal distribution is in general regularly varying, i.e., X_0 is a regularly varying random variable (Theorem 3.3.4). The same statement holds true for GARCH(p,q), see Basrak et al. [5]. As regards A.3, we recall that a stationary sequence $(\mathbf{Y}_t)$ is strongly mixing if

$$\sup_{\substack{A\in\sigma(\mathbf{Y}_t,\, t\leq 0)\\ B\in\sigma(\mathbf{Y}_t,\, t>k)}} |\,\mathbb{P}(A\cap B) - \mathbb{P}(A)\,\mathbb{P}(B)\,| =: a_k \to 0, \qquad k\to\infty.$$

If $(a_k)_{k\in\mathbb{N}}$ decays to zero at a geometric rate, then $(\mathbf{Y}_t)$ is said to be strongly mixing with geometric rate; see Ibragimov and Linnik [68], Bradley [24], Doukhan [40] or Rio [118]. Theorem 3.2 in Mikosch and Straumann [103] is as follows.

Theorem 7.1.1. *Consider the martingale transform*

$$(\sum_{t=1}^n \mathbf{Y}_t)_{n\in\mathbb{N}} = (\sum_{t=1}^n \mathbf{H}_t Y_t)_{n\in\mathbb{N}}$$

defined above. Assume that the conditions A.1 – A.3 *are satisfied. Moreover, if* $\kappa \in (1,2)$ *assume that* $\mathbb{E}Y_0 = 0$ *and, if* $\kappa = 1$*, that* Y_0 *is symmetric. Then*

$$a_n^{-1}\mathbf{S}_n \xrightarrow{\text{d}} \mathbf{D}_\kappa, \tag{7.2}$$

where the sequence (a_n) *is given by*

$$\mathbb{P}(|Y_0| > a_n) \sim n^{-1}, \qquad n\to\infty, \tag{7.3}$$

and $\mathbf{D}_\kappa$ *is a* κ*-stable random vector. If* $(\mathbf{Y}_t)$ *has extremal index* $\epsilon > 0$*, then the random vector* $\mathbf{D}_\kappa$ *is furthermore nondegenerate.*

Remark 7.1.2. We refrain from giving the definition of the extremal index and refer to Leadbetter, Lindgren and Rootzén [82] instead. It is worth mentioning that the extremal index of $(\mathbf{Y}_t)$ exists under A.1 – A.3; see Theorem 2.1

in Mikosch and Straumann [103]. κ–stable random vectors are usually defined in terms of their characteristic function. See Samorodnitsky and Taqqu [121] for definitions and properties of κ–stable distributions. Any linear combination of the components of an κ–stable random vector is again (univariate) κ–stable. A nondegenerate κ–stable random variable is regularly varying with index κ; see Feller [50]. Concerning the asymptotic behavior of $(a_n)_{n\geq 1}$, it is a well–known fact from the theory of regular variation that (7.3) together with the regular variation of Y_0 implies

$$a_n = \tilde{L}(n)\, n^{1/\kappa}, \qquad n \to \infty,$$

for some slowly varying $\tilde{L}$; see Bingham et al. [14], or Resnick [117]. From this observation together with (7.2) we read off that the rate of convergence of $\mathbf{S}_n/n$ in Theorem 7.1.1 is

$$n a_n^{-1} = (\tilde{L}(n))^{-1}\, n^{1-1/\kappa}.$$

Since $\tilde{L}(n)/n^s \to 0$ and $\tilde{L}(n)n^s \to \infty$ for all $s > 0$, the function $\tilde{L}_n$ is "flat" compared to a power function and plays the role of a constant. From $\kappa \in (0,2)$, we conclude that the rate of convergence of $\mathbf{S}_n/n$ is slower than the $\sqrt{n}$ rate in the central limit theorem for finite variance random variables. □

The proof of Theorem 7.1.1 rests on a general limit theorem for sums of dependent and heavy–tailed random vectors, see Theorem 2.8 in Davis and Mikosch [35]. One of the conditions, which have to be verified in order that Theorem 2.8 can be applied, is the (multivariate) regular variation of the finite–dimensional distributions of $(\mathbf{Y}_t)$; see Resnick [116, 117] for a definition of multivariate regular variation of random vectors. The regular variation property of $(\mathbf{Y}_t)$ follows by a famous result of Breiman [27]: if $X, Y > 0$ are two independent random variables, X is regularly varying with index $\kappa > 0$ and $\mathbb{E}Y^\kappa < \infty$, then

$$\mathbb{P}(XY > z) \sim \mathbb{E}Y^\kappa\, \mathbb{P}(X > z), \qquad z \to \infty, \tag{7.4}$$

i.e., the product XY is regularly varying with the same index κ. The conditions A.1 – A.2 and the special structure of $\mathbf{Y}_t = \mathbf{H}_t Y_t$ together with an application of (7.4) show that $|\mathbf{Y}_t|$ is regularly varying with index κ. With refined arguments, and repeated application of property (7.4), one can establish the regular variation of the finite–dimensional distributions of $(\mathbf{Y}_t)$. For detailed background information we refer to Mikosch and Straumann [103].

7.2 Infinite Variance Stable Limits of the QMLE

This section treats the limit behavior of the QMLE in the general conditionally heteroscedastic time series model (5.1) when $\mathbb{E}Z_0^4 = \infty$. This complements Theorem 5.6.1. We keep the notation of Chapter 5. Before we start, we modify the condition N.3 of Section 5.6 by omitting $\mathbb{E}Z_0^4 < \infty$:

N.3' The following moment conditions hold:

(i) $\mathbb{E}\left(\frac{|h_0'(\boldsymbol{\theta}_0)|^2}{\sigma_0^4}\right) < \infty,$

(ii) $\mathbb{E}\|\ell_0''\|_K < \infty.$

We recall that

$$\hat{L}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \hat{\ell}_t(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{n}\left(\frac{X_t^2}{\hat{h}_t(\boldsymbol{\theta})} + \log \hat{h}_t(\boldsymbol{\theta})\right)$$

and

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \ell_t(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{n}\left(\frac{X_t^2}{h_t(\boldsymbol{\theta})} + \log h_t(\boldsymbol{\theta})\right).$$

The QMLE $\hat{\boldsymbol{\theta}}_n$ is defined as a maximizer of $\hat{L}_n$ with respect to $\boldsymbol{\theta} \in K$, where $K \subset \Theta \subset \mathbb{R}^d$ is compact, and the sequence $\tilde{\boldsymbol{\theta}}_n$ maximizes the function L_n. First we identify the limit determining term for $\hat{\boldsymbol{\theta}}_n$ and $\tilde{\boldsymbol{\theta}}_n$.

Proposition 7.2.1. *Let* $((X_t, \sigma_t))$ *be the stationary ergodic solution in model* (5.1) *with true parameter vector* $\boldsymbol{\theta}_0$. *Suppose the conditions* N.1,N.2, N.3' *and* N.4 *hold true. Then the* QMLE $\hat{\boldsymbol{\theta}}_n$ *is strongly consistent. If there is a positive sequence* $(x_n)_{n\geq 1}$ *with* $x_n = o(n)$ *as* $n \to \infty$ *and*

$$x_n \frac{L_n'(\boldsymbol{\theta}_0)}{n} \xrightarrow{\text{d}} \mathbf{D}, \qquad n \to \infty, \tag{7.5}$$

for an $\mathbb{R}^d$*–valued random variable* $\mathbf{D}$, *then the* QMLE $\hat{\boldsymbol{\theta}}_n$ *satisfies the limit relation*

$$x_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\text{d}} -\mathbf{F}_0^{-1}\mathbf{D}, \qquad n \to \infty, \tag{7.6}$$

where the matrix $\mathbf{F}_0 = \mathbb{E}[\ell_0''(\boldsymbol{\theta}_0)]$ *is negative definite.*

Proof. The arguments are identical to the ones used in the proof of Theorem 5.6.1. We repeat them here for the sake of completeness. Since in the proofs for the strong consistency of $\hat{\boldsymbol{\theta}}_n$ (Theorem 5.3.1) and the negative definiteness of $\mathbf{F}_0$ (Lemma 5.6.3) we did not use that $\mathbb{E}Z_0^4 < \infty$, the QMLE $\hat{\boldsymbol{\theta}}_n$ is strongly consistent and the matrix $\mathbf{F}_0$ is negative definite under the assumptions of the present proposition. Analogously to the derivation of the limit relation (5.65), one first shows that the minimizer $\tilde{\boldsymbol{\theta}}_n$ of L_n obeys

$$x_n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\mathbf{F}_0^{-1}(1 + o_{\mathbb{P}}(1))\, x_n \frac{L_n'(\boldsymbol{\theta}_0)}{n}, \qquad n \to \infty.$$

This together with (7.5) and a Slutsky argument gives

$$x_n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\text{d}} -\mathbf{F}_0^{-1}\mathbf{D}, \qquad n \to \infty.$$

In a second step one establishes

$$x_n(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n) \xrightarrow{\text{a.s.}} \mathbf{0}, \qquad n \to \infty. \tag{7.7}$$

This and an application of Slutsky's lemma then imply the limit result (7.5). Relation (7.7) is derived as follows. Use the same arguments as in the proof of Lemma 5.6.5 in order to show

$$x_n(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n) = \mathbf{F}_0^{-1}(1 + o_{\mathbb{P}}(1))\,\frac{x_n}{n}\,(\hat{L}'_n(\hat{\boldsymbol{\theta}}_n) - L'_n(\hat{\boldsymbol{\theta}}_n)), \qquad n \to \infty. \tag{7.8}$$

The identical steps as employed in Lemma 5.6.4 yield $\|\hat{L}'_n - L'_n\|_K \le \sum_{t=1}^{\infty} \|\hat{\ell}'_t - \ell'_t\|_K < \infty$. From this bound applied to (7.8) and $\lim_{n\to\infty} x_n/n = 0$ we deduce the relation (7.7). Thus we have completed the proof. □

Now we state the main result.

Theorem 7.2.2. *Suppose that* $((X_t, \sigma_t))$ *is the stationary ergodic solution to model* (5.1) *with* $\boldsymbol{\theta}_0$. *Suppose the conditions* N.1,N.2, N.3' *and* N.4 *hold true. In addition, assume:*

(i) *The sequence* $\left(\dfrac{h'_t(\boldsymbol{\theta}_0)}{2\sigma_t^2}\,(Z_t^2 - 1)\right)$ *is strongly mixing with a geometric rate.*

(ii) *The random variable* Z_0^2 *is regularly varying with index* $\kappa \in (1, 2)$.

Then the QMLE $\hat{\boldsymbol{\theta}}_n$ *is strongly consistent and*

$$x_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\text{d}} \mathbf{D}_\kappa, \qquad n \to \infty,$$

where $\mathbf{D}_\kappa$ *is* κ*–stable and* $(x_n)_{n\ge 1} = (n a_n^{-1})_{n \ge 1}$ *with* $(a_n)_{n\ge1}$ *given by*

$$\mathbb{P}(Z_0^2 > a_n) \sim n^{-1}, \qquad n \to \infty.$$

If the stationary sequence $\left(\frac{h'_t(\boldsymbol{\theta}_0)}{\sigma_t^2}\,(Z_t^2 - 1)\right)$ *has an extremal index different from zero, then the random vector* $\mathbf{D}_\kappa$ *is nondegenerate.*

Remarks 7.2.3. Before proving the theorem, we discuss its practical consequences for parameter inference:

1. The rate of convergence x_n has — roughly speaking — magnitude $n^{1-1/\kappa}$, which is less than $\sqrt{n}$. The heavier the tails of the innovations, i.e., the smaller κ, the slower is the convergence of $\hat{\boldsymbol{\theta}}_n$ towards the true parameter $\boldsymbol{\theta}_0$.

2. The limit distribution of the standardized differences $(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is κ–stable and hence non–Gaussian. The exact parameters of this κ–stable limit are not explicitly known.

3. Confidence bands based on the normal approximation of Theorem 5.6.1 are false if $\mathbb{E}Z_0^4 = \infty$. □

Proof of Theorem 7.2.2. The proof follows from a combination of Proposition 7.2.1 and Theorem 7.1.1. One can apply Theorem 7.1.1 with $\mathbf{H}_t = h'_t(\boldsymbol{\theta}_0)/(2\sigma_t^2)$ and $Y_t = (Z_t^2 - 1)$ to

$$L'_n(\boldsymbol{\theta}_0) = \frac{1}{2}\sum_{t=1}^{n} \frac{h'_t(\boldsymbol{\theta}_0)}{\sigma_t^2}(Z_t^2 - 1).$$

We verify A.1 – A.2 of the aforementioned theorem. One can show that $\mathbb{P}(Y_0 > z) = \mathbb{P}(Z_0^2 - 1 > z) \sim \mathbb{P}(Z_0^2 > z)$ as $z \to \infty$ (use e.g. Theorem A3.2 in Embrechts et al. [45]), and thus $Y_0 = Z_0^2 - 1$ is also regularly varying with index κ and $\mathbb{P}(Y_0 > a_n) \sim n^{-1}$ as $n \to \infty$. Thus we have verified A.1. Assumption A.2 is fulfilled by virtue of N.3' (i) together with $\kappa \in (1,2)$. Thus from the application of Theorem 7.1.1,

$$a_n^{-1} L'_n(\boldsymbol{\theta}_0) = x_n \frac{L'_n(\boldsymbol{\theta}_0)}{n} \xrightarrow{\mathrm{d}} \tilde{\mathbf{D}}_\kappa,$$

where $\tilde{\mathbf{D}}_\kappa$ is κ–stable. Proposition 7.2.1 implies

$$x_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} -\mathbf{F}_0^{-1}\tilde{\mathbf{D}}_\kappa = \mathbf{D}_\kappa.$$

Recalling that a linear transform of an κ–stable random vector is again κ–stable shows that $\mathbf{D}_\kappa$ is κ–stable. When the sequence $\left(\frac{h'_t(\boldsymbol{\theta}_0)}{\sigma_t^2}(Z_t^2 - 1)\right)$ has nonzero extremal index, then $\tilde{\mathbf{D}}_\kappa$ is nondegenerate, and since $\mathbf{F}_0^{-1}$ is regular, also $\mathbf{D}_\kappa$ is nondegenerate. □

7.3 Limit Behavior of the QMLE in GARCH(p,q) with Heavy–tailed Innovations

The main difficulty in the application of Theorem 7.2.2 to GARCH(p,q) is the verification of the mixing property (i). The demonstration of strong mixing of a sequence of dependent random variables is often a nontrivial task. At the time being we can e.g. not deliver a proof for strong mixing of the sequence $(\mathbf{Y}_t)$ in AGARCH(p,q). We state the theorem and remind once again that Hall and Yao [63] derived the identical result by means of different techniques. Berkes and Horváth [6] confine themselves to a derivation of the rate of convergence of the QMLE in GARCH.

Theorem 7.3.1. *Let (X_t) be a stationary* GARCH(p,q) *process with true parameter vector $\boldsymbol{\theta}_0 = (\alpha_0^\circ, \ldots, \alpha_p^\circ, \beta_1^\circ, \ldots, \beta_q^\circ)^T$. Suppose that Z_0^2 is regularly varying with index $\kappa \in (1,2)$ and that* Q.1 – Q.4 *of Section 4.2.1 hold true. Moreover, assume that Z_0 has a Lebesgue density f, where the closure of the interior of $\{f > 0\}$ contains the origin. Define $(x_n)_{n\geq 1} = (na_n^{-1})_{n\geq 1}$, where*

$$P(Z_0^2 > a_n) \sim n^{-1}, \qquad n \to \infty.$$

Then the QMLE $\hat{\boldsymbol{\theta}}_n$ *is consistent and*

$$x_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} \mathbf{D}_\kappa, \qquad n \to \infty, \tag{7.9}$$

for some nondegenerate κ*–stable random vector* $\mathbf{D}_\kappa$*.*

Remark 7.3.2. By the definition of a GARCH process, the distribution of the innovations Z_t is unknown. Therefore assumptions about the heaviness of the tails of its distribution are purely hypothetical. Depending on the assumptions on the distribution of Z_0, one can develop different asymptotic theories for QMLE of GARCH processes: asymptotic normality as provided by Theorem 4.2.1 or infinite variance stable distribution as provided by Theorem 7.3.1.

In view of this rather disturbing result, one could think of introducing an alternative estimator for $\boldsymbol{\theta}_0$, whose asymptotic properties do not break down in case of infinite fourth moment innovations. Horváth and Liese [66] solved this problem for the ARCH(p) process. They show that the so–called weighted L_1 estimator does not require any moment condition on Z_0 for asymptotic normality. □

Proof of Theorem 7.3.1. The proof is an application of Theorem 7.2.2. Under Q.1 – Q.4 the conditions N.1, N.2 and N.4 hold true as can be seen from an evident adaptation of the arguments in Sections 5.4.2 and 5.7.1; these arguments did not rely on $\mathbb{E}Z_0^4 < \infty$. Some care has to be taken with respect to the moment conditions of N.3'. By Lemma 5.2 in Berkes et al. [8] the random variable $\|h_0'/h_0\|_K$ has finite moments of any order and thus N.3' (i) is true; we also appeal to Lemma 5.7.5 of this monograph. Concerning, N.3' (ii), Lemma 5.1 in Berkes et al. [8] (cf. Lemma 5.7.4) says that $\mathbb{E}\|X_0^2/h_0\|_K^\nu < \infty$ for every $\nu \in (0, \kappa)$. This fact together with an application of the Minkowski inequality to the decomposition (5.73) yields $\mathbb{E}\|\ell_0''\|_K < \infty$ so that N.3 (ii) is established. Consequently we are left to prove that the stationary sequence

$$(\mathbf{Y}_t) = \left(\frac{h_t'(\boldsymbol{\theta}_0)}{2\sigma_t^2} \left(Z_t^2 - 1 \right) \right),$$

is strongly mixing with geometric rate and that it has extremal index $\epsilon > 0$. Since the proof of the strong mixing with geometric rate is a problem on its own and rather lengthy, it is deferred to Section 7.4. As regards $\epsilon > 0$, according to Theorem 3.7.2 in Leadbetter et al. [82], if $\epsilon = 0$, then if for some sequence $(u_n)_{n\geq 1}$ the relation $\liminf_{n\to\infty} \mathbb{P}(\tilde{M}_n \leq u_n) > 0$ holds, one necessarily has $\lim_{n\to\infty} \mathbb{P}(M_n \leq u_n) = 1$. Here $M_n = \max(|\mathbf{Y}_1|, \ldots, |\mathbf{Y}_n|)$ and $(\tilde{M}_n)$ is the corresponding sequence of partial maxima for an iid sequence (R_t) where R_0 has the same distribution as $|\mathbf{Y}_0|$. Assume $\epsilon = 0$ by contradiction. The random variable $|\mathbf{Y}_0|$ is regularly varying with index κ by Breiman's result (7.4) together with N.3' (i). Hence $(a_n^{-1}\tilde{M}_n)_{n\geq 1}$ converges weakly to a Fréchet distribution, see e.g. Embrechts et al. [45], Chapter 3. More precisely, with $\Phi_\kappa(x) = e^{-x^{-\kappa}}$ one has $\mathbb{P}(\tilde{M}_n \leq a_n x) \to \Phi_\kappa(\zeta x)$ as $n \to \infty$ for all $x > 0$, where

$\zeta = (\mathbb{E}|h_0'(\boldsymbol{\theta}_0)/(2\sigma_0^2)_0|^\kappa)^{1/\kappa}$. Thus $\liminf_{n\to\infty} \mathbb{P}(\tilde{M}_n \le a_n x) = \Phi_\kappa(\zeta x) > 0$ for all $x > 0$. However, $\mathbb{P}(M_n \le x a_n) \to 1$ does not hold for all positive x. Indeed, make use of (5.43) with $\gamma = 0$ to obtain the a.s. representation

$$h_t(\boldsymbol{\theta}) = \pi_0(\boldsymbol{\theta}) + \sum_{\ell=1}^{\infty} \pi_\ell(\boldsymbol{\theta}) X_{t-\ell}^2, \qquad \boldsymbol{\theta} \in K,$$

where $\pi_0(\boldsymbol{\theta}) = \alpha_0/\beta_{\boldsymbol{\theta}}(1)$ and $\sum_{\ell=1}^{\infty} \pi_\ell(\boldsymbol{\theta}) z^\ell = \alpha_{\boldsymbol{\theta}}(z)/\boldsymbol{\theta}_{\boldsymbol{\theta}}(z)$, $|z| \le 1$, with $\alpha_{\boldsymbol{\theta}}(z) = \sum_{i=1}^{p} \alpha_i z^i$ and $\beta_{\boldsymbol{\theta}}(z) = 1 - \sum_{j=1}^{q} \beta_j z^j$. As shown in the proof of Lemma 5.4.3, there is $r > 0$ such that the polynomials $\beta_{\boldsymbol{\theta}}(z)$ have no roots in the disc $\{|z| \le 1 + r\}$ for *all* $\boldsymbol{\theta} \in K$, and hence the above Taylor series $\sum_{\ell=1}^{\infty} \pi_\ell(\boldsymbol{\theta}) z^\ell$ converges for $|z| \le 1 + r$. Now straightforward arguments exploiting

$$\sum_{\ell=1}^{\infty} \frac{\partial \pi_\ell(\boldsymbol{\theta})}{\partial \alpha_i} z^\ell = \frac{z^i}{\beta_{\boldsymbol{\theta}}(z)}, \qquad |z| \le 1 + r,$$

for all $i = 0, \ldots, p$ and $\sum_{j=1}^{q} \beta_j < 1$ on K show that

$$\frac{\partial \pi_\ell(\boldsymbol{\theta})}{\partial \alpha_i} \ge 0 \quad \text{and} \quad \sum_{\ell=1}^{\infty} \Big(\sum_{i=0}^{p} \alpha_i \frac{\partial \pi_\ell(\boldsymbol{\theta})}{\partial \alpha_i} \Big) z^\ell = \frac{\alpha_{\boldsymbol{\theta}}(z)}{\beta_{\boldsymbol{\theta}}(z)}, \qquad |z| \le 1 + r,$$

for all $\boldsymbol{\theta} \in K$. Thus the partial derivatives of h_t satisfy

$$\frac{\partial h_t(\boldsymbol{\theta})}{\partial \alpha_i} \ge 0 \quad \text{for all } i = 0, \ldots, p \quad \text{and} \quad \sum_{i=0}^{p} \alpha_i \frac{\partial h_t(\boldsymbol{\theta})}{\partial \alpha_i} = h_t(\boldsymbol{\theta}). \tag{7.10}$$

Since the Euclidean norm is equivalent to the 1–norm $|\mathbf{y}| = \sum_{\ell=1}^{p+q+1} |y_\ell|$ and $0 \le \alpha_i, \beta_j \le M$ on K for a certain constant $M > 0$, there is $c > 0$ such that

$$\begin{aligned} \frac{|h_t'(\boldsymbol{\theta})|}{h_t(\boldsymbol{\theta})} &\ge \frac{c}{h_t(\boldsymbol{\theta})} \left(\sum_{i=0}^{p} \alpha_i \left| \frac{\partial h_t(\boldsymbol{\theta})}{\partial \alpha_i} \right| + \sum_{j=1}^{q} \beta_j \left| \frac{\partial h_t(\boldsymbol{\theta})}{\partial \beta_j} \right| \right) \\ &\ge \frac{c}{h_t(\boldsymbol{\theta})} \sum_{i=0}^{p} \alpha_i \left| \frac{\partial h_t(\boldsymbol{\theta})}{\partial \alpha_i} \right| = \frac{c}{h_t(\boldsymbol{\theta})} \sum_{i=0}^{p} \alpha_i \frac{\partial h_t(\boldsymbol{\theta})}{\partial \alpha_i} = c. \end{aligned} \tag{7.11}$$

Note that the last two equalities in the latter display are a consequence of (7.10). Then by $\sigma_t^2 = h_t(\boldsymbol{\theta}_0)$ and inequality (7.11),

$$|\mathbf{Y}_t| = \frac{|h_t'(\boldsymbol{\theta}_0)|}{2\sigma_t^2} |Z_t^2 - 1| \ge \frac{c}{2} |Z_t^2 - 1|, \qquad t \in \mathbb{Z}.$$

This and $\mathbb{P}(|Z_0^2 - 1| > a_n) \sim \mathbb{P}(Z_0^2 > a_n) \sim n^{-1}$ as $n \to \infty$ imply

$$\mathbb{P}(M_n \le x a_n) \le \mathbb{P}\Big(\max_{t \le n} |Z_t^2 - 1| \le 2c^{-1} a_n x \Big) \to \Phi_\alpha(2c^{-1}x) < 1, \qquad n \to \infty.$$

From this contradiction we conclude $\epsilon > 0$, and this finishes the proof. □

Before we proceed by the verification of the mixing property, we illustrate Theorem 7.3.1 by means of a simulation. We choose GARCH(1, 1) parameters

$$\alpha_0^\circ = 1.492 \times 10^{-6},\ \alpha_1^\circ = 0.06706,\ \beta_1^\circ = 0.9079 \tag{7.12}$$

and suppose $Z_t \sim t_\nu$, where t_ν is the Student t distribution with ν degrees of freedom standardized to unit variance. We compare two cases. When $\nu = 5$ the moment condition $\mathbb{E}Z_0^4 < \infty$ in Theorem 4.2.1 holds true and the QMLE is asymptotically normal. In the case $\nu = 3$ the random variable Z_0^2 is regularly varying with index 3/2 and hence the regular variation condition of Theorem 7.3.1 applies. Then the QMLE converges at rate $n^{1-2/3} = n^{1/3}$. The boxplots in Figure 7.1 below are based on 500 independent replicates of $\hat\alpha_1$ for various sample sizes n. It is clearly visible that the speed of convergence to the true value α_1° is slower when $\nu = 3$. Furthermore we illustrate in Figure 7.2 that the normal approximation (according to Theorem 4.2.1) is misleading when $\nu = 3$.

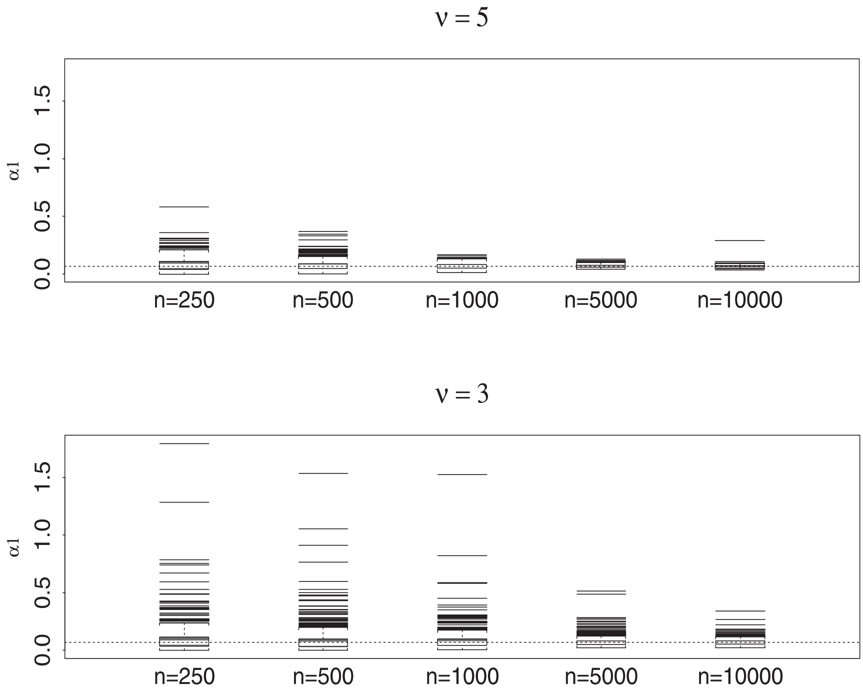

Fig. 7.1. Boxplots of independent realizations of the QMLE of $\hat\alpha_1$ in the GARCH(1, 1) process with parameters (7.12) and Student t_ν innovations. The dotted horizontal lines represent the true value α_1°.

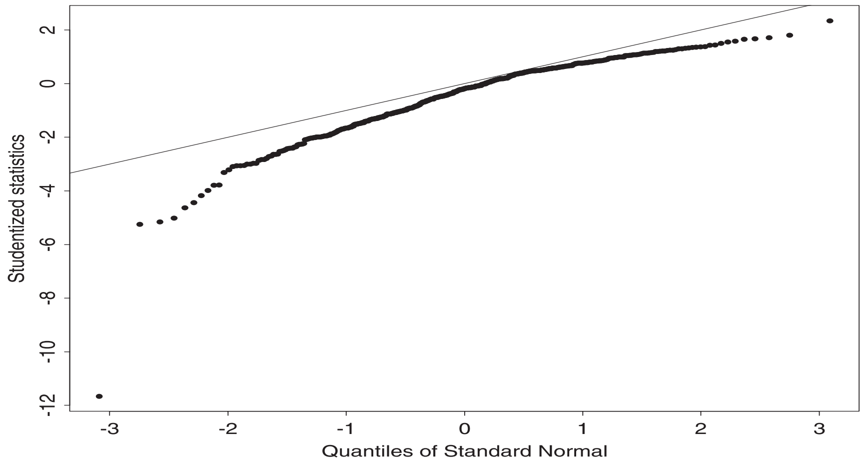

Fig. 7.2. GARCH(1, 1) model with parameters (7.12) and Student t_3 innovations. QQ–plot of 500 independent realizations of the studentized statistics of $\hat{\alpha}_1$ for sample size $n = 10000$. Moreover the line $\{x = y\}$ is drawn.

7.4 Verification of Strong Mixing with Geometric Rate of $(\mathbf{Y}_t)$ in GARCH(p,q)

To begin with, we quote a powerful result due to Mokkadem [104], which allows one to establish strong mixing in stationary solutions of so–called polynomial linear stochastic recurrence equations (SREs). A sequence $(\mathbf{Y}_t)$ of random vectors in $\mathbb{R}^d$ obeys a linear SRE if

$$\mathbf{Y}_{t+1} = \mathbf{P}_t \mathbf{Y}_t + \mathbf{Q}_t, \qquad t \in \mathbb{Z}, \tag{7.13}$$

where $((\mathbf{P}_t, \mathbf{Q}_t))$ constitutes an iid sequence with values in $\mathbb{R}^{d \times d} \times \mathbb{R}^d$. A linear SRE is called polynomial if there exists an iid sequence $(\mathbf{e}_t)$ in $\mathbb{R}^{d'}$ such that $\mathbf{P}_t = \mathbf{P}(\mathbf{e}_t)$ and $\mathbf{Q}_t = \mathbf{Q}(\mathbf{e}_t)$, where $\mathbf{P}(\mathbf{x})$ and $\mathbf{Q}(\mathbf{x})$ have entries and coordinates, respectively, which are polynomial functions of the coordinates of $\mathbf{x}$. The existence and uniqueness of a stationarity solution to (7.13) has been studied by Brandt [25], Bougerol and Picard [22], Babillot et al. [3] and others. The following set of conditions is sufficient, compare with the proof of Theorem 3.3.1: $\mathbb{E}(\log^+ \|\mathbf{P}_0\|_{\mathrm{op}}) < \infty$, $\mathbb{E}(\log^+ |\mathbf{Q}_0|) < \infty$, and the top Lyapunov exponent associated with the matrix operator sequence $(\mathbf{P}_t)$ is strictly negative, i.e.,

$$\rho = \inf_{t \in \mathbb{N}} \left\{ \frac{1}{t+1} \, \mathbb{E}\big(\log \|\mathbf{P}_0 \cdots \mathbf{P}_{-t}\|_{\mathrm{op}} \big) \right\} < 0. \tag{7.14}$$

Here $\|\cdot\|_{\mathrm{op}}$ is the matrix operator norm corresponding to the Euclidean norm (2.9). We also need the notion of absolute regularity (or β–mixing), which is more restrictive than strong mixing. A stationary sequence $(\mathbf{Y}_t)$ is absolutely regular if

$$\mathbb{E}\left(\sup_{B\in\sigma(\mathbf{Y}_t,\, t>k)} |\mathbb{P}(B \mid \sigma(\mathbf{Y}_t,\, t \le 0)\,) - \mathbb{P}(B)|\right) =: \; b_k \to 0, \qquad k \to \infty. \tag{7.15}$$

Indeed, absolute regularity implies strong mixing with the same mixing coefficients b_k. The following result is an easy generalization of Theorem 4.3 in Mokkadem [104].

Theorem 7.4.1. *Let* $(\mathbf{e}_t)$ *be an iid sequence of random vectors in* $\mathbb{R}^{d'}$. *Then consider the polynomial linear* SRE

$$\mathbf{Y}_{t+1} = \mathbf{P}(\mathbf{e}_t)\mathbf{Y}_t + \mathbf{Q}(\mathbf{e}_t), \qquad t \in \mathbb{Z}, \tag{7.16}$$

where $\mathbf{P}(\mathbf{e}_t)$ *is a random* $d \times d$ *matrix and* $\mathbf{Q}(\mathbf{e}_t)$ *a random vector in* $\mathbb{R}^d$. *Suppose:*

(1) $\mathbf{P}(\mathbf{0})$ *has spectral radius strictly smaller than* 1 *and the top Lyapunov exponent* ρ *corresponding to* $(\mathbf{P}(\mathbf{e}_t))$ *is strictly negative.*

(2) *There is* $s > 0$ *such that*

$$\mathbb{E}||\mathbf{P}(\mathbf{e}_0)||_{op}^s < \infty \quad \text{and} \quad \mathbb{E}|\mathbf{Q}(\mathbf{e}_0)|^s < \infty.$$

(3) *There is a smooth algebraic variety* $V \subset \mathbb{R}^{d'}$ *such that* $\mathbf{e}_0$ *has a density* f *with respect to Lebesguian measure on* V. *Assume that* $\mathbf{0}$ *is contained in the closure of the interior of the set* $\{f > 0\}$.

Then the polynomial linear SRE (7.16) *has a unique stationary solution* $(\mathbf{Y}_t)$, *which is ergodic and absolutely regular with geometric rate and consequently strongly mixing with geometric rate.*

Remark 7.4.2. As regards the definition of a smooth algebraic variety, we first introduce the notion of an algebraic subset. An algebraic subset of $\mathbb{R}^{d'}$ is a set of form

$$V = \{\mathbf{x} \in \mathbb{R}^{d'} \mid F_1(\mathbf{x}) = \cdots = F_r(\mathbf{x}) = \mathbf{0}\},$$

where $F_1, \ldots, F_r$ are real multivariate polynomials. An algebraic variety is an algebraic subset which is not the union of two proper algebraic subsets. An algebraic variety is smooth if the Jacobian of $\mathbf{F} = (F_1, \ldots, F_r)^T$ has identical rank everywhere on V. Examples of smooth algebraic varieties in $\mathbb{R}^{d'}$ are the hyperplanes of $\mathbb{R}^{d'}$ or $V = \mathbb{R}^{d'}$. □

Remarks 7.4.3. The original Theorem 4.3 by Mokkadem [104] differs from the above theorem only in a stronger moment condition for $||\mathbf{P}(\mathbf{e}_0)||_{\text{op}}$. Instead of condition (2), Mokkadem imposes:

(2)' There is $s > 0$ such that $\mathbb{E}||\mathbf{P}(\mathbf{e}_0)||_{\text{op}}^s < 1$ and $\mathbb{E}|\mathbf{Q}(\mathbf{e}_0)|^s < \infty$.

Note also that (2)' makes the requirement of a strictly negative top Lyapunov exponent of $(\mathbf{P}(\mathbf{e}_t))$ in (1) obsolete since by Jensen's inequality

$$\mathbb{E}(\log \|\mathbf{P}(\mathbf{e}_0)\|_{\text{op}}) = s^{-1}\mathbb{E}(\log \|\mathbf{P}(\mathbf{e}_0)\|_{\text{op}}^s) \leq s^{-1}\log(\mathbb{E}\|\mathbf{P}(\mathbf{e}_0)\|_{\text{op}}^s) < 0.$$

Unfortunately the moment condition $\mathbb{E}\|\mathbf{P}(\mathbf{e}_0)\|_{\text{op}}^s < 1$ is often not easy to check or, even worse, not fulfilled. This is e.g. the case for GARCH, but the same difficulty appears also in bilinear time series models, which have a Markovian representation of form (7.16). Several authors devised highly sophisticated criteria to circumvent this problem, see e.g. Feigin and Tweedie [49], Pham [113] or Tjøstheim [128]. Our generalization in Theorem 7.4.1 offers simpler remedy.

The proof of Theorem 4.3 in Mokkadem [104] makes use of techniques from general state space Markov chain theory and algebraic topology and is by no means simple. The special structure of $\mathbf{P}(\mathbf{e}_t)$ and $\mathbf{Q}(\mathbf{e}_t)$ and conditions (1)–(3) enable one to construct a σ–finite measure μ such that the Markov chain started at some initial value at time $t = 0$ and obeying (7.16) is μ–irreducible; then Markov chain theory can be applied. Mokkadem formulates his results in the framework of so–called polynomial autoregressive processes, a more general class than the polynomial linear SREs. We mention that Carrasco and Chen [31] apply Theorem 4.3 of Mokkadem [104] to several conditionally heteroscedastic time series models and that Boussama [23] uses the theory of Mokkadem [104] in order to establish absolute regularity with geometric rate of GARCH(p, q). □

Proof of Theorem 7.4.1. There is nothing to prove if $\mathbb{E}\|\mathbf{P}(\mathbf{e}_0)\|_{\text{op}}^{\tilde{s}} < 1$ for some $\tilde{s} > 0$ as this special case is the content of Theorem 4.3 in Mokkadem [104]. For the general case it suffices to prove the absolute regularity with geometric rate for some subsequence $(\mathbf{Y}_{tm})_{t\in\mathbb{Z}}$, where $m \geq 1$ is fixed. Indeed, the mixing coefficient b_k is nonincreasing and since $(\mathbf{Y}_t)$ is a Markov process, the simpler representation

$$b_k = \mathbb{E}\left(\sup_{B\in\sigma(\mathbf{Y}_{k+1})} |\,\mathbb{P}(B \mid \sigma(\mathbf{Y}_0)) - \mathbb{P}(B)|\right)$$

is also valid, see e.g. Bradley [24]. Since $\rho < 0$ and $(\mathbf{P}_t)$ is stationary, there is $m \geq 1$ with $\mathbb{E}(\log \|\mathbf{P}(\mathbf{e}_m)\cdots\mathbf{P}(\mathbf{e}_1)\|_{\text{op}}) < 0$. From the fact that the first derivative of the map $u \mapsto \mathbb{E}\|\mathbf{P}(\mathbf{e}_m)\cdots\mathbf{P}(\mathbf{e}_1)\|_{\text{op}}^u$ equals $\mathbb{E}(\log \|\mathbf{P}(\mathbf{e}_m)\cdots\mathbf{P}(\mathbf{e}_1)\|_{\text{op}}) <$ 0 at $u = 0$, we deduce that there is $0 < \tilde{s} \leq s$ with $\mathbb{E}\|\mathbf{P}(\mathbf{e}_m)\cdots\mathbf{P}(\mathbf{e}_1)\|_{\text{op}}^{\tilde{s}} < 1$. Then note that $(\tilde{\mathbf{Y}}_t) = (\mathbf{Y}_{tm})$ obeys a polynomial linear SRE:

$$\tilde{\mathbf{Y}}_{t+1} = \tilde{\mathbf{P}}(\tilde{\mathbf{e}}_t)\tilde{\mathbf{Y}}_t + \tilde{\mathbf{Q}}(\tilde{\mathbf{e}}_t), \qquad t \in \mathbb{Z},$$

where $$\tilde{\mathbf{e}}_t = \begin{pmatrix} \mathbf{e}_{(t+1)m-1} \\ \vdots \\ \mathbf{e}_{tm} \end{pmatrix}$$

and

$$\tilde{\mathbf{P}}(\tilde{\mathbf{e}}_t) = \mathbf{P}(\mathbf{e}_{(t+1)m-1}) \cdots \mathbf{P}(\mathbf{e}_{tm}),$$

$$\tilde{\mathbf{Q}}(\tilde{\mathbf{e}}_t) = \mathbf{Q}(\mathbf{e}_{(t+1)m-1}) + \sum_{\ell=1}^{m-1} \left(\prod_{i=1}^{\ell} \mathbf{P}(\mathbf{e}_{(t+1)m-i}) \right) \mathbf{Q}(\mathbf{e}_{(t+1)m-\ell-1}).$$

Since both the matrix $\tilde{\mathbf{P}}(\tilde{\mathbf{e}}_t)$ and the vector $\tilde{\mathbf{Q}}(\tilde{\mathbf{e}}_t)$ are polynomial functions of the coordinates of $\tilde{\mathbf{e}}_t$ and the sequence $(\tilde{\mathbf{e}}_t)$ is iid, $(\tilde{\mathbf{Y}}_t)$ obeys a polynomial linear SRE. Observe that $\tilde{\mathbf{P}}(\mathbf{0}) = (\mathbf{P}(\mathbf{0}))^m$ has spectral radius strictly smaller than 1, that $\mathbb{E}\|\tilde{\mathbf{P}}(\tilde{\mathbf{e}}_0)\|_{\text{op}}^{\tilde{s}} < 1$ and $\mathbb{E}|\tilde{\mathbf{Q}}(\tilde{\mathbf{e}}_0)|^{\tilde{s}} < \infty$ and that $\tilde{\mathbf{e}}_0$ has a density with respect to Lebesgue measure on V^m, where V^m is a smooth algebraic variety (see A.14 in Mokkadem [104]). This density obeys (3). Thus an application of Theorem 4.3 in Mokkadem [104] yields that $(\tilde{\mathbf{Y}}_t)$ is absolutely regular with geometric rate. This proves the assertion. □

We now give an auxiliary result.

Lemma 7.4.4. *Let* $(\mathbf{P}_t)$ *be an iid sequence of* $k \times k$*–matrices with* $\mathbb{E}\|\mathbf{P}_0\|_{op}^{s} < \infty$ *for some* $s > 0$*. Then the associated top Lyapunov exponent* $\rho < 0$ *if and only if there exist* $c > 0$, $\tilde{s} > 0$ *and* $0 \le \lambda < 1$ *so that*

$$\mathbb{E}\|\mathbf{P}_t \cdots \mathbf{P}_1\|_{op}^{\tilde{s}} \le c\lambda^t, \qquad t \ge 1. \tag{7.17}$$

Proof. For the proof of necessity, observe that by definition (7.14) of the top Lyapunov exponent there exists $m \ge 1$ such that $\mathbb{E}(\log \|\mathbf{P}_m \cdots \mathbf{P}_1\|_{\text{op}}) < 0$. From the fact that the map $u \mapsto \mathbb{E}\|\mathbf{P}_m \cdots \mathbf{P}_1\|_{\text{op}}^{u}$ has first derivative equal to $\mathbb{E}(\log \|\mathbf{P}_m \cdots \mathbf{P}_1\|_{\text{op}})$ at $u = 0$ and $(\mathbf{P}_t)$ is stationary, we deduce that there is $\tilde{s} > 0$ with $\mathbb{E}\|\mathbf{P}_m \cdots \mathbf{P}_1\|_{\text{op}}^{\tilde{s}} = \tilde{\lambda} < 1$. Since the matrix operator norm $\|\cdot\|_{\text{op}}$ is submultiplicative and the factors in $\mathbf{P}_t \cdots \mathbf{P}_1$ are iid,

$$\mathbb{E}\|\mathbf{P}_t \cdots \mathbf{P}_1\|^{\tilde{s}} \le \tilde{\lambda}^{[t/m]} \left(\max_{\ell=1,\dots,m-1} \mathbb{E}\|\mathbf{P}_\ell \cdots \mathbf{P}_1\|_{\text{op}}^{\tilde{s}} \right) \le c\lambda^t, \qquad t \ge 1,$$

for $c = \tilde{\lambda}^{-1} \left(\max_{\ell=1,\dots,m-1} \mathbb{E}\|\mathbf{P}_\ell \cdots \mathbf{P}_1\|^{\tilde{s}} \right)$ and $\lambda = \tilde{\lambda}^{1/m}$. Regarding the proof of sufficiency, the definition (7.14) together with Jensen's inequality leads to

$$\begin{aligned}
\rho &\le \frac{1}{t}\, \mathbb{E}(\log \|\mathbf{P}_t \cdots \mathbf{P}_1\|_{\text{op}}) = \frac{1}{t\tilde{s}}\, \mathbb{E}(\log \|\mathbf{P}_t \cdots \mathbf{P}_1\|_{\text{op}}^{\tilde{s}}) \\
&\le \frac{1}{t\tilde{s}} \log(\mathbb{E}\|\mathbf{P}_t \cdots \mathbf{P}_1\|_{\text{op}}^{\tilde{s}}) \le \frac{1}{t\tilde{s}} (\log c + t \log \lambda) \to \frac{\log \lambda}{\tilde{s}}, \qquad t \to \infty,
\end{aligned}$$

and shows $\rho < 0$. This completes the proof of the lemma. □

The next proposition shows that the top Lyapunov exponent of block lower triangular matrices is strictly negative if and only if the building blocks have a strictly negative top Lyapunov exponent. The same statement is of course true for block *upper* triangular matrices.

Proposition 7.4.5. *Suppose that*

$$\mathbf{P}_t = \begin{pmatrix} \mathbf{A}_t & \mathbf{0}_{r\times(k-r)} \\ \mathbf{B}_t & \mathbf{C}_t \end{pmatrix}, \qquad t \in \mathbb{Z}, \tag{7.18}$$

forms an iid sequence of $k \times k$–matrices with $\mathbb{E}\|\mathbf{P}_0\|_{op}^s < \infty$ for some $s > 0$, where $\mathbf{A}_t \in \mathbb{R}^{r\times r}$, $\mathbf{B}_t \in \mathbb{R}^{(k-r)\times r}$ and $\mathbf{C}_t \in \mathbb{R}^{(k-r)\times(k-r)}$. Then its associated top Lyapunov exponent $\rho_{\mathbf{P}} < 0$ if and only if the sequences $(\mathbf{A}_t)$ and $(\mathbf{C}_t)$ have top Lyapunov exponents $\rho_{\mathbf{A}} < 0$ and $\rho_{\mathbf{C}} < 0$.

Proof. For the proof of sufficiency of $\rho_{\mathbf{A}} < 0$ and $\rho_{\mathbf{C}} < 0$ for $\rho_{\mathbf{P}} < 0$, it is by Lemma 7.4.4 enough to derive a moment inequality of form (7.17) for $(\mathbf{P}_t)$. Since all matrix norms are equivalent, we may without loss of generality replace $\|\cdot\|_{\text{op}}$ by the Frobenius norm (2.7), denoted by $\|\cdot\|$. By induction we obtain

$$\mathbf{P}_t \cdots \mathbf{P}_1 = \begin{pmatrix} \mathbf{A}_t \cdots \mathbf{A}_1 & \mathbf{0}_{r\times(k-r)} \\ \mathbf{Q}_t & \mathbf{C}_t \cdots \mathbf{C}_1 \end{pmatrix},$$

where

$$\begin{aligned} \mathbf{Q}_t &= \mathbf{B}_t\mathbf{A}_{t-1}\cdots\mathbf{A}_1 + \mathbf{C}_t\mathbf{B}_{t-1}\mathbf{A}_{t-2}\cdots\mathbf{A}_1 + \mathbf{C}_t\mathbf{C}_{t-1}\mathbf{B}_{t-2}\mathbf{A}_{t-3}\cdots\mathbf{A}_1 \\ &\quad + \cdots + \mathbf{C}_t\cdots\mathbf{C}_3\mathbf{B}_2\mathbf{A}_1 + \mathbf{C}_t\cdots\mathbf{C}_2\mathbf{B}_1. \end{aligned}$$

Since we use the Frobenius norm, the following bounds are valid:

$$\begin{aligned} \max\left(\|\mathbf{A}_t\cdots\mathbf{A}_1\|, \|\mathbf{C}_t\cdots\mathbf{C}_1\|\right) &\le \|\mathbf{P}_t\cdots\mathbf{P}_1\| \\ &\le \|\mathbf{A}_t\cdots\mathbf{A}_1\| + \|\mathbf{C}_t\cdots\mathbf{C}_1\| + \|\mathbf{Q}_t\|. \end{aligned} \tag{7.19}$$

It is sufficient to show (7.17) (with $\|\cdot\|_{\text{op}}$ replaced by $\|\cdot\|$) for each block in the matrix $\mathbf{P}_t\cdots\mathbf{P}_1$. Because of $\rho_{\mathbf{A}} < 0$, $\rho_{\mathbf{C}} < 0$ and $\mathbb{E}\|\mathbf{A}_0\|^s, \mathbb{E}\|\mathbf{C}_0\|^s \le \mathbb{E}\|\mathbf{P}_0\|^s < \infty$, Lemma 7.4.4 already implies moment bounds of form (7.17) for $(\mathbf{A}_t)$ and $(\mathbf{C}_t)$. Thus we are left to bound $\|\mathbf{Q}_t\|$. Without loss of generality we may assume that the constants $\lambda < 1$ and $\tilde{s}, c > 0$ in the inequality (7.17) are equal for $(\mathbf{A}_t)$ and $(\mathbf{C}_t)$ and that $\tilde{s} \le s \le 1$. From an application of the Minkowski inequality and exploiting the independence of the factors in each summand of $\mathbf{Q}_t$, we receive the desired relation

$$\mathbb{E}\|\mathbf{Q}_t\|^{\tilde{s}} \le c^2 t\, \mathbb{E}\|\mathbf{B}_0\|^{\tilde{s}}\lambda^{t-1} \le \tilde{c}\,\tilde{\lambda}^t,$$

some $\tilde{\lambda} \in (\lambda, 1)$, $\tilde{c} > 0$. For the proof of necessity, assume $\rho_{\mathbf{P}} < 0$. Then the left–hand estimates in (7.19) and Lemma 7.4.4 imply that $\rho_{\mathbf{A}} < 0$ and $\rho_{\mathbf{C}} < 0$. This completes the proof. □

Finally we can exploit Theorem 7.4.1 and establish strong mixing with geometric rate of the sequence $(\mathbf{Y}_t) = (h_t'(\boldsymbol{\theta}_0)(Z_t^2-1)/(2\sigma_t^2))$ in GARCH(p,q).

Proposition 7.4.6. *Under the conditions of Theorem 7.3.1 the sequence* $(\mathbf{Y}_t)$ *is absolutely regular with geometric rate.*

Proof. For the proof of this result we first embed $(\mathbf{Y}_t)$ in a polynomial linear SRE. Without loss of generality assume $p, q \geq 3$. Write

$$\tilde{\mathbf{Y}}_t = \Big(\sigma_t^2, \ldots, \sigma_{t-q+1}^2, X_{t-1}^2, \ldots, X_{t-p+1}^2,$$
$$\frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial \alpha_0}, \ldots, \frac{\partial h_{t-q+1}(\boldsymbol{\theta}_0)}{\partial \alpha_0}, \ldots, \frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial \alpha_p}, \ldots, \frac{\partial h_{t-q+1}(\boldsymbol{\theta}_0)}{\partial \alpha_p},$$
$$\frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial \beta_1}, \ldots, \frac{\partial h_{t-q+1}(\boldsymbol{\theta}_0)}{\partial \beta_1}, \ldots, \frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial \beta_q}, \ldots, \frac{\partial h_{t-q+1}(\boldsymbol{\theta}_0)}{\partial \beta_q}\Big)^T.$$

Since $Z_t^2 = X_t^2/\sigma_t^2$, we have

$$\sigma(\mathbf{Y}_t,\, t > k) \subset \sigma(\tilde{\mathbf{Y}}_{t+1}, t > k) \quad \text{and} \quad \sigma(\mathbf{Y}_t,\, t \leq 0) \subset \sigma(\tilde{\mathbf{Y}}_{t+1},\, t \leq 0).$$

Consequently, it is enough to demonstrate absolute regularity with geometric rate of the sequence $(\tilde{\mathbf{Y}}_t)$. The goal is to derive a linear polynomial SRE for $(\mathbf{Y}_t)$. To this end we introduce various matrices. Write $\mathbf{0}_{d_1 \times d_2}$ for the $d_1 \times d_2$ matrix with all entries equal to zero and let $\mathbf{I}_d$ denote the identity matrix of dimension d. Then set

$$\mathbf{M}_1(Z_t) = \begin{pmatrix} \boldsymbol{\tau}_t & \beta_q^\circ & \boldsymbol{\alpha}^\circ & \alpha_p^\circ \\ \mathbf{I}_{q-1} & \mathbf{0}_{(q-1)\times 1} & \mathbf{0}_{(q-1)\times(p-2)} & \mathbf{0}_{(q-1)\times 1} \\ \boldsymbol{\zeta}_t & \mathbf{0}_{1\times 1} & \mathbf{0}_{1\times(p-2)} & \mathbf{0}_{1\times 1} \\ \mathbf{0}_{(p-2)\times(q-1)} & \mathbf{0}_{(p-2)\times 1} & \mathbf{I}_{p-2} & \mathbf{0}_{(p-2)\times 1} \end{pmatrix},$$

where

$$\boldsymbol{\tau}_t = (\beta_1^\circ + \alpha_1^\circ Z_t^2, \beta_2^\circ, \ldots, \beta_{q-1}^\circ) \in \mathbb{R}^{1\times(q-1)},$$
$$\boldsymbol{\zeta}_t = (Z_t^2, 0, \ldots, 0) \in \mathbb{R}^{1\times(q-1)},$$
$$\boldsymbol{\alpha}^\circ = (\alpha_2^\circ, \ldots, \alpha_{p-1}^\circ) \in \mathbb{R}^{1\times(p-2)}.$$

Note that $\mathbf{M}_1(Z_t)$ coincides with (3.10) with $\gamma = 0$. Moreover, define

$$\mathbf{M}_2(Z_t) = \begin{pmatrix} \mathbf{0}_{q\times(p+q-1)} \\ \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_p \end{pmatrix} \qquad \text{and} \qquad \mathbf{M}_4 = \begin{pmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_q \end{pmatrix},$$

where $\mathbf{U}_i \in \mathbb{R}^{q\times(p+q-1)}$ and $\mathbf{V}_j \in \mathbb{R}^{q\times(p+q-1)}$ are given by

$$[\mathbf{U}_1]_{k,\ell} = \delta_{k\ell,11}\, Z_t^2,$$

$$[\mathbf{U}_i]_{k,\ell} = \delta_{k\ell,1(q+i-1)}, \qquad i \geq 2,$$

$$[\mathbf{V}_j]_{k,\ell} = \delta_{k\ell,1j}.$$

Here, $\delta_{\cdot}$ denotes the Kronecker symbol. Also introduce the $q \times q$ matrix

$$\mathbf{C} = \begin{pmatrix} \beta_1^\circ & \cdots & \beta_q^\circ \\ \mathbf{I}_{q-1} & & \mathbf{0}_{(q-1)\times 1} \end{pmatrix},$$

and let

$$\mathbf{M}_3 = \operatorname{diag}(\mathbf{C}, p+1), \qquad \mathbf{M}_5 = \operatorname{diag}(\mathbf{C}, q)$$

be the block diagonal matrices consisting of $p+1$ (or q) copies of the block $\mathbf{C}$. Finally, we define

$$\mathbf{P}(Z_t) = \begin{pmatrix} \mathbf{M}_1(Z_t) & \mathbf{0}_{(p+q-1)\times(p+1)q} & \mathbf{0}_{(p+q-1)\times q^2} \\ \mathbf{M}_2(Z_t) & \mathbf{M}_3 & \mathbf{0}_{(p+1)q\times q^2} \\ \mathbf{M}_4 & \mathbf{0}_{q^2\times(p+1)q} & \mathbf{M}_5 \end{pmatrix}$$

and $\mathbf{Q} \in \mathbb{R}^{p+q-1+q(p+q+1)}$ by $[\mathbf{Q}]_k = \alpha_0\delta_{k,1} + \delta_{k,p+q}$. Differentiating both sides of $h_{t+1}(\boldsymbol{\theta}) = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t+1-i}^2 + \sum_{j=1}^q \beta_j h_{t+1-j}(\boldsymbol{\theta})$ at the true parameter $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we recognize that

$$h'_{t+1}(\boldsymbol{\theta}_0) =$$
$$(1, X_t^2, \ldots, X_{t-p+1}^2, \sigma_t^2, \ldots, \sigma_{t-q+1}^2) + \beta_1^\circ h'_t(\boldsymbol{\theta}_0) + \cdots + \beta_q^\circ h'_{t+1-q}(\boldsymbol{\theta}_0).$$

From this recursive relationship together with $\sigma_{t+1}^2 = \alpha_0^\circ + \alpha_1^\circ X_t^2 + \cdots + \alpha_p^\circ X_{t-p+1}^2 + \beta_1^\circ \sigma_t^2 + \cdots + \beta_q^\circ \sigma_{t-q+1}^2$ we have that $(\tilde{\mathbf{Y}}_t)$ obeys the following polynomial linear SRE on $\mathbb{R}^{p+q-1+q(p+q+1)}$:

$$\mathbf{y}_{t+1} = \mathbf{P}(Z_t)\mathbf{y}_t + \mathbf{Q}. \tag{7.20}$$

The proof of Proposition 7.4.6 follows if we can show that the polynomial liner SRE (7.20) obeys the conditions of Theorem 7.4.1. Then its unique stationary solution, $(\tilde{\mathbf{Y}}_t)$, is ergodic and absolutely regular with geometric rate. The verification is carried out in the following lemma □

Lemma 7.4.7. *Under the assumptions of Proposition 7.4.6, the polynomial linear* SRE (7.20) *obeys the conditions of Theorem 7.4.1.*

Proof. Since $\mathbb{E}Z_0^2 = 1$, it is immediate that $\mathbb{E}\|\mathbf{P}(Z_0)\|_{\text{op}} < \infty$ since this statement is true for the Frobenius norm and all matrix norms are equivalent. Treat the blocks $\mathbf{M}_1(Z_t)$, $\mathbf{M}_3$ and $\mathbf{M}_4$ separately. Recall that the matrix

$\mathbf{M}_1(Z_t)$ appeared in Section 3.3.1 as the random transition map for the vectors $(\sigma_t^2, \ldots, \sigma_{t-q+1}^2, X_{t-1}^2, \ldots X_{t-p+1}^2)^T$ in GARCH(p,q); choose $\gamma = 0$ in (3.10). Theorem 3.3.1 states that stationarity of GARCH(p,q) is equivalent to $(\mathbf{M}_1(Z_t))$ having a strictly negative top Lyapunov exponent. Moreover, arguing by recursion on p and expanding the determinant with respect to the last column, it is easily verified that $\mathbf{M}_1(0)$ has characteristic polynomial

$$\det(\lambda \mathbf{I}_{p+q-1} - \mathbf{M}_1(0)) = \lambda^{p+q-1}\left(1 - \sum_{j=1}^{q} \beta_j^\circ \lambda^{-j}\right).$$

Since $\sum_{j=1}^{q} \beta_j^\circ < 1$ in a stationary GARCH(p,q) process, by repeated application of the triangle inequality

$$\left|1 - \sum_{j=1}^{q} \beta_j^\circ \lambda^{-j}\right| \geq 1 - \sum_{j=1}^{q} \beta_j^\circ \lambda^{-j} \geq 1 - \sum_{j=1}^{q} \beta_j^\circ > 0$$

if $|\lambda| \geq 1$, and hence $\mathbf{M}_1(0)$ has spectral radius < 1. Observe that the building block $\mathbf{C}$ has characteristic polynomial

$$\det(\lambda \mathbf{I}_q - \mathbf{C}) = \lambda^q \left(1 - \sum_{j=1}^{q} \beta_j^\circ \lambda^{-j}\right),$$

showing that its spectral radius is strictly smaller than 1 (use the same argument as before). Thus the *deterministic* matrices $\mathbf{M}_3$ and $\mathbf{M}_5$ have spectral radius < 1, which also implies that their associated top Lyapunov exponents are strictly negative. Combining these results, we deduce that $\mathbf{P}(0)$ has spectral radius < 1 and conclude by twice applying Proposition 7.4.5 that $(\mathbf{P}(Z_t))$ has strictly negative top Lyapunov exponent. Condition (3) of Theorem 7.4.1 is automatically fulfilled since we assumed in Theorem 7.3.1 that Z_0 has a Lebesgue density f, where the closure of the interior of $\{f > 0\}$ contains the origin. This concludes the proof of the lemma. □

Remark 7.4.8. Since (X_t^2, σ_t^2) is a subvector of $\tilde{\mathbf{Y}}_{t+1}$, stationary GARCH(p,q) processes are absolutely regular with geometric rate; this result has previously been established by Boussama [23]. □

8

Whittle Estimation in a Heavy–tailed GARCH(1,1) Model

This chapter is mainly based on the paper by Mikosch and Straumann [102] and can be seen as a continuation of the discussion in Section 4.2.2. It consists of a detailed study of the limit properties of the Whittle estimator applied to the squares of a stationary GARCH(1,1) process (X_t). Thereby we extend the results obtained by Giraitis and Robinson [56], which are summarized by Theorem 4.2.3 of this monograph: the Whittle estimator is strongly consistent and asymptotically normal provided the process has a marginal distribution with a finite 8th moment.

In this chapter we focus on the case when $\mathbb{E}X_0^8 = \infty$. This case corresponds to various real–life log–return series of financial data. We show that the Whittle estimator is consistent as long as the 4th moment is finite and inconsistent when the 4th moment is infinite. Moreover, in the finite 4th moment case, the rates of convergence of the Whittle estimator to the true parameter are the slower the fatter the tail of the distribution of X_0. These findings are in contrast to Whittle estimation of ARMA processes with iid innovations. Indeed, in the latter case it was shown in Mikosch et al. [98] that the rate of convergence of the Whittle estimator to the true parameter is the faster the fatter the tails of the innovations distributions.

8.1 Introduction

We maintain the notation of the previous chapters. To make this chapter as self–contained as possible, we briefly summarize the most important facts about GARCH(1,1) and recall the definition of the Whittle estimator.

GARCH(1,1)

We consider a GARCH(1,1) process (X_t), which is given by the equations

$$X_t = \sigma_t Z_t \quad \text{and} \quad \sigma_{t+1}^2 = \alpha_0 + \alpha_1 X_t^2 + \beta_1 \sigma_t^2, \qquad t \in \mathbb{Z}, \qquad (8.1)$$

where (Z_t) is a sequence of iid *symmetric* random variables with $\mathrm{Var}(Z_0) = 1$, (σ_t) is a nonnegative process, and $\alpha_0 > 0$, $\alpha_1, \beta_1 \geq 0$. Observe that, in contrast to Chapter 3, we have added the *technical* assumption that Z_0 is symmetric, i.e.,

$$Z_0 \stackrel{\mathrm{d}}{=} -Z_0 \,.$$

This restriction has been made for the proof of Theorem 8.2.1 below; it is not clear whether this condition can be weakened. The sequence (σ_t^2) obeys the linear SRE

$$\sigma_{t+1}^2 = A_t \sigma_t^2 + B_t, \qquad t \in \mathbb{Z}, \tag{8.2}$$

where $A_t = \alpha_1 Z_t^2 + \beta_1$, $B_t = \alpha_0$, and $((A_t, B_t))$ iid. Since the latter SRE has a unique strictly stationary solution if and only if

$$\mathbb{E}(\log A_0) = \mathbb{E}[\,\log(\alpha_1 Z_0^2 + \beta_1)\,] < 0, \tag{8.3}$$

condition (8.3) is sufficient and necessary for a unique stationary GARCH$(1,1)$ process (8.1) to exist (Nelson [105]). In particular, $\beta_1 < 1$ is necessary for (8.3) since

$$0 > \mathbb{E}[\,\log(\alpha_1 Z_0^2 + \beta_1)\,] \geq \log(\beta_1)\,. \tag{8.4}$$

In what follows, we always assume that condition (8.3) is satisfied and that (X_t) is a strictly stationary GARCH$(1,1)$ process. We have discussed in Section 3.3.1 that the marginal distributions of (A)GARCH have heavy tails. For this chapter we assume the conditions of Theorem 3.3.4 with $\gamma = 0$ so that $|X_0|$ has a Pareto–like tail, i.e.,

$$\mathbb{P}(|X_0| > x) \sim \mathbb{E}|Z_0|^\kappa\, \mathbb{P}(\sigma_0 > x) \sim c_0\, x^{-\kappa}, \qquad x \to \infty, \tag{8.5}$$

for some $c_0 > 0$; recall that the tail index $\kappa > 0$ is related to the random variable A_0 through the equation

$$\mathbb{E}A_0^{\kappa/2} = 1. \tag{8.6}$$

Whittle Estimation

As shown in Section 4.2.2, every squared stationary GARCH$(1,1)$ process can be embedded in an ARMA(1,1) model:

$$X_t^2 = \varphi_1 X_{t-1}^2 + \nu_t - \beta_1 \nu_{t-1}, \qquad t \in \mathbb{Z}, \tag{8.7}$$

where $\nu_t = \sigma_t^2(Z_t^2 - 1)$, $\varphi_1 = \alpha_1 + \beta_1$ and the stationary sequence (ν_t) constitutes white noise if $\mathrm{Var}(X_0^2) < \infty$. This property leads one to consider the Whittle estimator of the squared GARCH process with model parameter

$$\boldsymbol{\vartheta} = (\varphi_1, \beta_1)^T = (\alpha_1 + \beta_1, \beta_1)^T.$$

Provided the variance of ν_0 is finite, the process $(X_t^2 - \mathbb{E}X_0^2)$ has spectral density

$$f(\lambda;\boldsymbol{\vartheta}) = \frac{\sigma_\nu^2}{2\pi}\, g(\lambda;\boldsymbol{\vartheta}), \qquad \lambda \in (-\pi,\pi],$$

where

$$g(\lambda;\boldsymbol{\vartheta}) = \frac{|1-\beta_1 e^{-i\lambda}|^2}{|1-\varphi_1 e^{-i\lambda}|^2}, \qquad \sigma_\nu^2 = \mathbb{E}\nu_0^2.$$

We learned from (8.4) that $\beta_1 < 1$ is a necessary condition for stationarity. Therefore we search for the minimum of the objective function

$$\bar{\sigma}^2_{n,X^2}(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_j \frac{I_{n,X^2}(\lambda_j)}{g(\lambda_j;\boldsymbol{\vartheta})} \tag{8.8}$$

(here $I_{n,X^2}(\lambda_j)$ denotes the periodogram at the Fourier frequency λ_j, the summation is taken over all Fourier frequencies (4.16)) on the set

$$C = \{\boldsymbol{\vartheta} \in \mathbb{R}^2 \,|\, 0 \le \beta_1 < 1,\ \beta_1 \le \varphi_1 \le 1\}. \tag{8.9}$$

The particular definition of the periodogram in (4.15) ensures that $I_{n,X^2}(0) = 0$ and therefore rules out irregular asymptotic behavior of the periodogram at zero. For $\lambda_j \neq 0$ the value of the periodogram $I_{n,X^2}(\lambda_j)$ is invariant with respect to the centering of the X_t^2's. It will turn out in the proofs below that centering of the X_t^2's becomes necessary when one wants to use the asymptotic results for the sample autocovariance function.

One observes that $\bar{\sigma}^2_{n,X^2}(\boldsymbol{\vartheta})$ has a minimum on the closure $\bar{C} = C \cup \{(1,1)\}$ of C. Therefore the following adaptation of the Whittle estimator is well defined:

$$\hat{\boldsymbol{\vartheta}}_n = \underset{\boldsymbol{\vartheta}\in\bar{C}}{\operatorname{argmin}}\ \bar{\sigma}^2_{n,X^2}(\boldsymbol{\vartheta})\,. \tag{8.10}$$

8.2 Limit Theory for the Sample Autocovariance Function

Recall that for any sample $Y_1,\dots,Y_n$ from a stationary sequence (Y_t), the sample autocovariance function (sample ACVF) is defined by

$$\gamma_{n,Y}(k) = \frac{1}{n}\sum_{t=1}^{n-|k|}(Y_t - \overline{Y})(Y_{t+|k|} - \overline{Y}), \qquad k \in \mathbb{Z}.$$

(for $|k| \ge n$ the sums are interpreted as zero), where $\bar{Y}$ denotes the sample mean, and the corresponding sample autocorrelation function (sample ACF) by

$$\rho_{n,Y}(k) = \frac{\gamma_{n,Y}(k)}{\gamma_{n,Y}(0)}, \qquad k \in \mathbb{Z}.$$

Their deterministic counterparts are the ACVF

$$\gamma_Y(k) = \mathrm{Cov}(Y_0, Y_k), \qquad k \in \mathbb{Z},$$

and the ACF

$$\rho_Y(k) = \gamma_Y(k)/\gamma_Y(0), \qquad k \in \mathbb{Z}.$$

In this section we formulate the basic asymptotic results for the sample ACF and sample ACVF of the squares of a stationary GARCH$(1,1)$ process (X_t). For its formulation we need the notions of stable random vector and multivariate stable distribution; we refer to the encyclopedic monograph by Samorodnitsky and Taqqu [121] for definitions and properties. The following results are given in Mikosch and Stărică [99].

Theorem 8.2.1. *Assume the conditions of Theorem 3.3.4 with $\gamma = 0$ hold and that the innovations (Z_t) are symmetric. Let (x_n) be a sequence of positive numbers given by*

$$x_n = \begin{cases} n^{1-4/\kappa}, & \text{if } \kappa < 8, \\ n^{1/2}, & \text{if } \kappa > 8, \end{cases} \qquad n \geq 1, \tag{8.11}$$

where κ is the tail index of $|X_0|$ as provided by Theorem 3.3.4. Then the following limit results hold.

(A) *The case $\kappa < 4$.*

$$x_n \left[\gamma_{n,X^2}(h) \right]_{h=0,\ldots,k} \xrightarrow{\mathrm{d}} (V_h)_{h=0,\ldots,k}\,, \tag{8.12}$$

$$\left[\rho_{n,X^2}(h) \right]_{h=1,\ldots,k} \xrightarrow{\mathrm{d}} (V_h/V_0)_{h=1,\ldots,k}\,, \tag{8.13}$$

where the vector $(V_0, \ldots, V_k)$ has positive components with probability one and it is jointly $\kappa/4$–stable in $\mathbb{R}^{k+1}$.

(B) *The case $4 < \kappa < 8$.*

$$x_n \left[\gamma_{n,X^2}(h) - \gamma_{X^2}(h) \right]_{h=0,\ldots,k} \xrightarrow{\mathrm{d}} (V_h)_{h=0,\ldots,k}\,, \tag{8.14}$$

$$x_n \left[\rho_{n,X^2}(h) - \rho_{X^2}(h) \right]_{h=1,\ldots,k} \xrightarrow{\mathrm{d}} \gamma_{X^2}^{-1}(0) \left[V_h - \rho_{X^2}(h) V_0 \right]_{h=1,\ldots,k}\,, \tag{8.15}$$

where the vector $(V_0, \ldots, V_k)$ is jointly $\kappa/4$–stable in $\mathbb{R}^{k+1}$.

(C) *The case $\kappa > 8$.*

$$x_n \left[\gamma_{n,X^2}(h) - \gamma_{X^2}(h) \right]_{h=0,\ldots,k} \xrightarrow{\mathrm{d}} (V_h)_{h=0,\ldots,k}\,, \tag{8.16}$$

$$x_n \left[\rho_{n,X^2}(h) - \rho_{X^2}(h) \right]_{h=1,\ldots,k} \xrightarrow{\mathrm{d}} \gamma_{X^2}^{-1}(0) \left[V_h - \rho_{X^2}(h) V_0 \right]_{h=1,\ldots,k}\,, \tag{8.17}$$

where the vector $(V_0, \ldots, V_k)$ is multivariate centered Gaussian.

Remark 8.2.2. An α–stable random variable Y with $\alpha < 2$ (the nondegenerate components of an α–stable random vector are α–stable as well) has tail $\mathbb{P}(|Y| > x) \sim cx^{-\alpha}$. Hence the limits of the sample ACVF in parts (A) and (B) have infinite variance distributions, in part (A) even infinite first moment limits.

In part (A), the ACF and ACVF of (X_t^2) are not defined since $EX_0^4 = \infty$. The sample ACF converges weakly to a distribution with finite support.

In part (B), the ACVF and ACF of (X_t^2) are well defined. In view of (8.11), the rate of convergence of the sample ACF to the ACF is the slower the closer κ to 4.

In part (C), X_0^2 has finite variance, and the limit results are a consequence of a standard CLT for strongly mixing sequences with geometric rate. □

Remark 8.2.3. In the case $\kappa < 4$, the sample ACVF and sample ACF can be replaced by the corresponding versions for the noncentered $X_1^2, \ldots, X_n^2$. This follows from the results in Davis and Mikosch [35]. A particular consequence is that the limiting random variables V_h are positive with probability 1. □

8.3 Main Results

Now we are ready to formulate the main results on the asymptotic behavior of the Whittle estimator for the squared GARCH$(1,1)$ case. We start with the consistency.

Theorem 8.3.1. *Let (X_t) be a strictly stationary* GARCH$(1,1)$ *process with parameter vector $\boldsymbol{\vartheta}_0 = (\varphi_1^\circ, \beta_1^\circ)^T = (\alpha_1^\circ + \beta_1^\circ, \beta_1^\circ)^T \in C$, satisfying the conditions of Theorem 3.3.4 with $\gamma = 0$ and symmetric innovations (Z_t). Then the following statements hold.*

(A)*If the tail index $\kappa < 4$ and $\alpha_1^\circ, \beta_1^\circ > 0$, i.e., $\boldsymbol{\vartheta}_0$ lies in the interior of C, the Whittle estimator $\hat{\boldsymbol{\vartheta}}_n$ defined in* (8.10) *is not consistent.*

(B)*If $\kappa > 4$, the Whittle estimator is strongly consistent.*

The inconsistency is illustrated in Figure 8.1. Part (B) of the theorem raises the question as to the rate of convergence of $\hat{\boldsymbol{\vartheta}}_n$ to $\boldsymbol{\vartheta}_0$. Here is the answer. But first recall the definition of (x_n) from (8.11):

$$x_n = \begin{cases} n^{1-4/\kappa}, & \text{if } \kappa < 8, \\ n^{1/2}, & \text{if } \kappa > 8. \end{cases}$$

Theorem 8.3.2. *In addition to the conditions of Theorem 8.3.1 assume that the tail index $\kappa > 4$ and $\mathbb{E}Z_0^8 < \infty$. Then the following limit relation holds:*

$$x_n(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{\mathrm{d}} [\mathbf{W}(\boldsymbol{\vartheta}_0)]^{-1} \Big(\mathbf{f}_0(\boldsymbol{\vartheta}_0) V_0 + 2 \sum_{k=1}^{\infty} \mathbf{f}_k(\boldsymbol{\vartheta}_0) V_k \Big), \qquad (8.18)$$

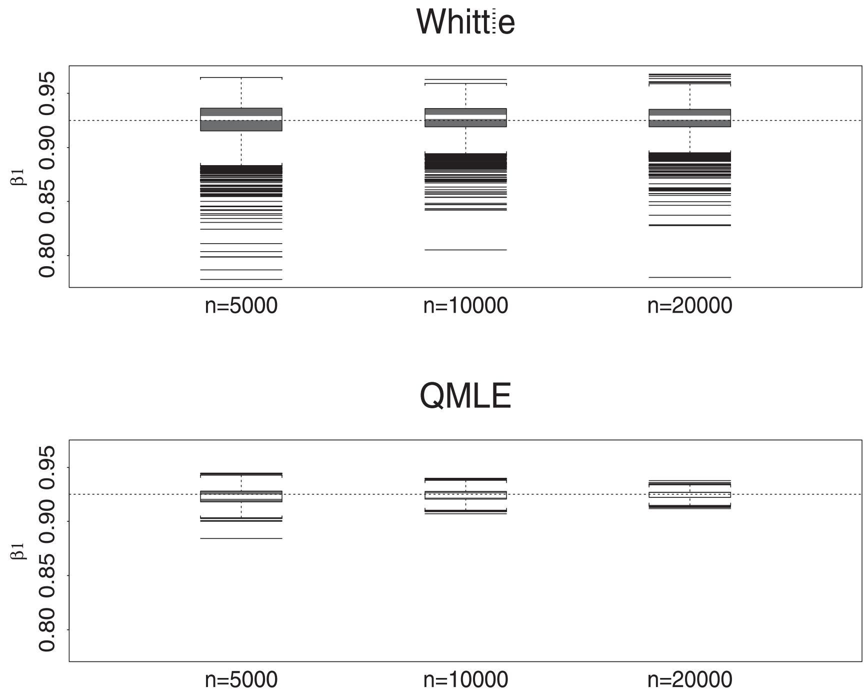

Fig. 8.1. GARCH(1,1) model with parameters $\alpha_0^\circ = 8.58 \times 10^{-6}$, $\alpha_1^\circ = 0.072$, $\beta_1^\circ = 0.925$ and standard Gaussian innovations. The tail index $\kappa = 3.2$ was determined by solving the equation (8.6) through Monte–Carlo simulation; see Mikosch [97] for more details. The finite sample distributions of the Whittle estimator of β_1, represented by boxplots of 2000 independent replicates (top), indicate that it is not consistent. In contrast, the QMLE of β_1 (bottom) converges at $\sqrt{n}$–rate. This is in line with Theorems 8.3.1 and 5.7.1.

where $(V_h)_{h=0,1,\ldots}$ *is a sequence of* $\kappa/4$*–stable random variables as specified in* (8.14) *for* $\kappa \in (4,8)$ *and a sequence of centered Gaussian random variables as specified in* (8.16) *for* $\kappa > 8$. *The infinite series on the right–hand side of* (8.18) *is understood as the weak limit of its partial sums. Moreover,* $[\mathbf{W}(\boldsymbol{\vartheta}_0)]^{-1}$ *is the inverse of the matrix*

$$\mathbf{W}(\boldsymbol{\vartheta}_0) = \frac{\sigma_\nu^2}{2\pi} \int_{-\pi}^{\pi} \left[\frac{\partial \log g(\lambda; \boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}}\right] \left[\frac{\partial \log g(\lambda; \boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}}\right]^T \mathrm{d}\lambda$$

and

$$\mathbf{f}_k(\boldsymbol{\vartheta}_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial(1/g(\lambda; \boldsymbol{\vartheta}_0))}{\partial \boldsymbol{\vartheta}} \, e^{-ik\lambda} \, \mathrm{d}\lambda, \qquad k \geq 0.$$

Remark 8.3.3. In the case of finite 8th moments, the above results follow from the paper of Giraitis and Robinson [56]. The proof of the corresponding part of Theorem 8.3.2 does not provide any additional work and so we included it for completeness. In contrast to the results in Giraitis and Robinson, we do not use martingale central limit theory. □

Remark 8.3.4. The rate of convergence of the QMLE in GARCH$(1,1)$ is determined by $L_n'(\boldsymbol{\theta}_0)$ (the function L_n was defined in (5.36)). Since

$$L_n'(\boldsymbol{\theta}_0) = -\frac{1}{2}\sum_{t=1}^{n} \frac{h_t'(\boldsymbol{\theta}_0)}{\sigma_t^2}\,(1 - Z_t^2)$$

and since $h_t'(\boldsymbol{\theta}_0)/h_t(\boldsymbol{\theta}_0)$ has moments of *any* order (see Berkes et al. [8]) and is independent of Z_t^2, the $\sqrt{n}$–rate and the asymptotic normality of the QMLE under $\mathbb{E}Z_0^4 < \infty$ of the QMLE (Theorem 4.2.1) is not totally surprising. In contrast, the Whittle estimator is virtually a function of the sample autocorrelations of (X_t^2) and their (slow) rate of convergence determines the convergence rate of the Whittle estimator. This may be seen as a key argument for the superiority of the QMLE over the Whittle estimator. □

Remark 8.3.5. As a matter of fact, the Whittle estimator is extremely flexible under various modifications of the ARMA model. For example, the Whittle estimator also works when estimating the parameters of an ARMA process with infinite variance innovations (Z_t) and can be extended to long memory FARIMA processes with or without infinite variance; see Mikosch et al. [98] for the ARMA case and Kokoszka and Taqqu [76] for the FARIMA case. It turns out that the $\sqrt{n}$–asymptotics for the Whittle estimator in the case of finite variance ARMA has to be replaced by more favorable rates of convergence in the infinite variance case. Roughly speaking, the Whittle estimator works the better the heavier the tails of the innovations (equivalently, the tails of the X_t's). A careful study of the proof shows that the results heavily depend on the faster than $\sqrt{n}$–rates of convergence for the sample ACVF and sample ACF of linear processes. These rates were derived by Davis and Resnick [36, 37]. Keeping in mind the slower than $\sqrt{n}$–rates of convergence for the sample ACF of the squared GARCH$(1,1)$ process when $\kappa \in (4,8)$, the rate $x_n = n^{1-4/\kappa}$ for the Whittle estimator $\hat{\boldsymbol{\vartheta}}_n$ is not totally unexpected. □

Remark 8.3.6. The gaps $\kappa = 4$ and $\kappa = 8$ in the above results are due to the fact that the corresponding results for the sample ACVF of the squared GARCH$(1,1)$ process are not yet available in these cases. □

Remark 8.3.7. (Compare with Remark 4.2.4(3)) In the above discussion we left out the estimation of the parameter α_0. Estimation of α_0 can be based on the formula

$$\mathrm{Var}(X_0) = \frac{\alpha_0^\circ}{1-\varphi_1^\circ}\,.$$

A natural estimator of α_0 is therefore given by

$$\hat{\alpha}_0 = \gamma_{n,X}(0)\,(1-\hat{\varphi}_1),$$

where $\hat{\varphi}_1$ is the Whittle estimator of φ. If $\kappa > 4$, then $\gamma_{n,X}(0) \xrightarrow{\text{a.s.}} \operatorname{Var}(X_0)$ by virtue of the ergodic theorem and, hence, $\hat{\alpha}_0$ is strongly consistent under the assumptions of Theorem 8.3.1. Moreover, under the conditions of Theorem 8.3.2,

$$\begin{aligned} x_n\,(\hat{\alpha}_0 - \alpha_0^\circ) &= (\gamma_{n,X}(0) - \gamma_X(0))\,x_n\,(1-\hat{\varphi}_1) + \gamma_X(0)\,x_n\,((-\hat{\varphi}_1 + \varphi_1^\circ)) \\ &\xrightarrow{\text{d}} \gamma_X(0)\,Y, \end{aligned}$$

where the limit distribution of Y is defined through Theorem 8.3.2. Therefore one can show via a continuous mapping argument that the joint limit distribution of $(x_n(\hat{\alpha}_0 - \alpha_0^\circ, \hat{\alpha}_1 - \alpha_1^\circ, \hat{\beta}_1 - \beta_1^\circ)^T)$ exists. It can again be expressed by the random variables (V_k). We omit details. □

8.4 Excursion: Yule–Walker Estimation in ARCH(p)

The Yule–Walker matrix equation for the AR(p) model $Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + Z_t$ for a white noise sequence (Z_t), $1 - \varphi_1 z - \cdots - \varphi_p z^p \neq 0$, $|z| \le 1$ (causality) is

$$\mathbf{R}\,\boldsymbol{\phi} = \boldsymbol{\rho}\,, \tag{8.19}$$

where $\mathbf{R}$ is the $p \times p$ matrix $(\rho_Y(i-j))_{i,j=1,\dots,p}$, $\boldsymbol{\phi} = (\varphi_1, \dots, \varphi_p)^T$ and $\boldsymbol{\rho} = (\rho_Y(1), \dots, \rho_Y(p))^T$, provided $\operatorname{Var}(Y_1) < \infty$. The Yule–Walker estimator of $\boldsymbol{\phi}$ is then obtained as the solution to (8.19) with $\mathbf{R}$ and $\boldsymbol{\rho}$ replaced by $\hat{\mathbf{R}}_n = (\rho_{n,Y}(i-j))_{i,j=1,\dots,p}$ and $\hat{\boldsymbol{\rho}}_n = (\rho_{n,Y}(1), \dots, \rho_{n,Y}(p))^T$, respectively. According to Brockwell and Davis [29], Proposition 5.1.1, $(\hat{\mathbf{R}}_n)^{-1}$ exists if $\gamma_{n,Y}(0) > 0$, and then

$$\hat{\boldsymbol{\phi}}_n = (\hat{\mathbf{R}}_n)^{-1}\,\hat{\boldsymbol{\rho}}_n. \tag{8.20}$$

From this representation it is immediate that $\hat{\boldsymbol{\phi}}_n$ estimates $\boldsymbol{\phi}$ consistently if the sample ACF is a consistent estimator of the ACF. Moreover, following the argument on p. 557 of Davis and Resnick [37], we conclude that

$$\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0 = \mathbf{D}\,(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) + o_{\mathbb{P}}(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_0) \tag{8.21}$$

for some nonsingular matrix $\mathbf{D}$.

Recall that the squares of an ARCH(p) process (X_t) can be written as an AR(p) process

$$X_t^2 = \alpha_0^\circ + \alpha_1^\circ X_{t-1}^2 + \cdots + \alpha_p X_{t-p}^2 + \nu_t\,, \tag{8.22}$$

where $\nu_t = X_t^2 - \sigma_t^2 = \sigma_t^2(Z_t^2 - 1)$ is a white noise sequence provided $\operatorname{Var}(X_0^2) < \infty$. If we replace in the above remarks (Y_t) by (X_t^2), the same arguments apply as long as the sample ACF of (X_t^2) is consistent. Thus the Yule–Walker estimator of the parameters α_i based on the AR(p) equation (8.22) is consistent. Assume that the statements of Theorems 3.3.4 with $\gamma = 0$

and 8.2.1 hold accordingly for ARCH(p) (see Basrak et al. [5] for a precise formulation). Then we also may conclude from (8.21) and Theorem 8.2.1 (with GARCH(1, 1) replaced by ARCH(p)) that in ARCH(p) the rate of convergence is the same as for the sample ACF:

$$x_n\,(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) \xrightarrow{\mathrm{d}} \mathbf{D}\,\mathbf{Y},$$

where for $\kappa > 4$, $\mathbf{Y} = \gamma_{X^2}^{-1}(0)(V_h - \rho_{X^2}(h)V_0)_{h=1,\ldots,p}$ with the specification of (V_h) as given in parts (B) and (C) of Theorem 8.2.1. For $\kappa < 4$, by virtue of part (A), a consistency result for the Yule–Walker estimator cannot be expected. Indeed, if $\gamma_{n,X^2}(0) > 0$, an appeal to (8.20) shows that the Yule–Walker estimator is a continuous function of the first p sample autocorrelations which converge weakly to a nondegenerate limit. For example, for $p = 1$ we obtain the usual estimator $\hat{\alpha}_1 = \rho_{n,X^2}(1)$ which has a nondegenerate limit distribution as described in part (A) of Theorem 8.2.1.

The Whittle estimator for an AR process is asymptotically equivalent to the Yule–Walker estimator and to the least–squares estimator defined in Section 4.1.2. (If one uses in the definition (8.8) an integral instead of a Riemann sum, the Yule–Walker and the Whittle estimator even coincide). Therefore its asymptotic properties only depend on a finite number of the sample autocorrelations and, therefore, an application of the continuous mapping theorem yields the limit distribution and convergence rate for the Yule–Walker estimator. The Whittle estimator based on the ARMA structure of a general squared GARCH(p, q) process is not as easily treated as the ARCH case since the Whittle estimator then depends on an increasing (with the sample size n) number of sample autocorrelations. This will become clear for the GARCH case in the proofs of Sections 8.5 and 8.6.

8.5 Proof of Theorem 8.3.1

The proof in the case $\kappa > 4$ is identical with the one for ARMA processes with iid noise as provided in Brockwell and Davis [29], Section 10.8. The proof only makes use of the ergodicity of (X_t); see Giraitis and Robinson [56] or Mikosch and Stărică [101, 100]. In the remainder of this section we study the case $\kappa < 4$.

For any compact $K \subset \mathbb{R}^2$, $\mathbb{C}(K)$ denotes the space of continuous functions on K equipped with the supremum topology and $\xrightarrow{\mathrm{d}}$ stands for convergence in distribution in $\mathbb{C}(K)$. Similarly, we write $\mathbb{C}(K, \mathbb{R}^2)$ for the space of 2-dimensional continuous functions on K, equipped with the supremum-topology. We use the same symbol $\xrightarrow{\mathrm{d}}$ for convergence in distribution in this space.

Proposition 8.5.1. *Assume the conditions of Theorem 8.3.1 hold and* $\kappa < 4$. *Then for any compact set* $K \subset C$*:*

(A)

$$u_n = \frac{\bar{\sigma}^2_{n,X^2}}{\gamma_{n,X^2}(0)} \xrightarrow{\mathrm{d}} v = \sum_{k\in\mathbb{Z}} \rho_k b_k \quad \textit{in } \mathbb{C}(K), \tag{8.23}$$

where $\rho_k = V_k/V_0$, (V_k) *is the sequence of limiting* $\kappa/4$*–stable random variables defined in* (8.12), *and*

$$b_k(\boldsymbol{\vartheta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{g(\lambda;\boldsymbol{\vartheta})}\, e^{-ik\lambda}\,\mathrm{d}\lambda.$$

(B)

$$\nabla u_n \xrightarrow{\mathrm{d}} \nabla v \quad \textit{in } \mathbb{C}(K,\mathbb{R}^2), \tag{8.24}$$

and the limiting process has representation

$$\nabla v = \sum_{k\in\mathbb{Z}} \rho_k\, \nabla b_k\,. \tag{8.25}$$

Here ∇ *denotes the gradient.*

Proof of Proposition 8.5.1. Part (A). We appeal to some of the ideas in the proof of Proposition 10.8.2 in Brockwell and Davis [29]. We start by observing that the Cesàro sum approximation

$$q_m(\lambda;\boldsymbol{\vartheta}) = \frac{1}{m}\sum_{j=0}^{m-1}\sum_{|k|\le j} b_k(\boldsymbol{\vartheta})e^{-ik\lambda} = \sum_{|k|<m}\left(1-\frac{|k|}{m}\right) b_k(\boldsymbol{\vartheta})e^{-ik\lambda}.$$

to the periodic function $q(\lambda;\boldsymbol{\vartheta}) = 1/g(\lambda;\boldsymbol{\vartheta})$ is uniform on $[-\pi,\pi]\times K$ (Theorem 2.11.1 in Brockwell and Davis [29]). Hence, for every $\epsilon>0$, there exists $m_0\ge 1$ such that for $m\ge m_0$,

$$\sup_{(\lambda,\boldsymbol{\vartheta})\in[-\pi,\pi]\times K} |q_m(\lambda;\boldsymbol{\vartheta}) - q(\lambda;\boldsymbol{\vartheta})| < \epsilon,$$

and as in (10.8.9) of Brockwell and Davis [29], for every $n\ge 1$,

$$\|u_n - u_{nm}\|_K \le \epsilon \quad \text{a.s.}, \tag{8.26}$$

where

$$u_n(\boldsymbol{\vartheta}) = \frac{1}{n\gamma_{n,X^2}(0)}\sum_j I_{n,X^2}(\lambda_j)/g(\lambda_j;\boldsymbol{\vartheta}),$$

$$u_{nm}(\boldsymbol{\vartheta}) = \frac{1}{n\gamma_{n,X^2}(0)}\sum_j I_{n,X^2}(\lambda_j)\, q_m(\lambda_j;\boldsymbol{\vartheta}).$$

We will show the following limit relations in $\mathbb{C}(K)$:

(a) For every fixed $m \geq 1$, as $n \to \infty$,

$$u_{nm} \xrightarrow{\mathrm{d}} v_m = \sum_{|k|<m} (1 - |k|/m)\, \rho_k\, b_k\,.$$

(b) As $m \to \infty$, $v_m \xrightarrow{\mathrm{d}} v$.

(c) For every $\epsilon > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\|u_{nm} - u_n\|_K > \epsilon\right) = 0\,, \tag{8.27}$$

where $\|\cdot\|_K$ denotes the supremum norm on K.

It then follows from Theorem 2.4.3 the desired relation

$$u_n \xrightarrow{\mathrm{d}} v \qquad \text{in } \mathbb{C}(K).$$

By virtue of (8.26), (c) is satisfied, and so it remains to prove (a) and (b). Before we proceed, we give two auxiliary results.

Lemma 8.5.2. *Under the conditions of Proposition 8.5.1, there exist* $0 < a < 1$ *and* $c > 0$ *such that*

$$|b_k(\boldsymbol{\vartheta})| \leq c a^{|k|}\,, \quad k \in \mathbb{Z},\, \boldsymbol{\vartheta} \in K.$$

Furthermore, the modulus of continuity of $b_k(\boldsymbol{\vartheta})$ *decays exponentially fast in* $|k|$*, i.e.,*

$$\sup_{|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|\leq\delta} |b_k(\boldsymbol{\vartheta}) - b_k(\boldsymbol{\vartheta}')| \leq c(\delta)\, a^{|k|},$$

where $\lim_{\delta\downarrow 0} c(\delta) = 0$*. The same statements remain valid if* b_k *is everywhere replaced by its gradient* ∇b_k*.*

The proof is standard by using that the function $f(z;\boldsymbol{\vartheta}) = (1-\varphi_1 z)/(1-\beta_1 z)$ is analytic with a power series representation with exponentially decaying coefficients, uniformly in $\boldsymbol{\vartheta} \in K$.

Lemma 8.5.3. *Under the conditions of Proposition 8.5.1, for each fixed* $k \geq 0$*,*

$$\rho_{n,X^2}(n-k) \xrightarrow{\mathbb{P}} 0, \qquad n \to \infty.$$

Proof of Lemma 8.5.3. Observe that

$$\begin{aligned}&\rho_{n,X^2}(n-k)\\ &= n^{-1}\left(\sum_{t=1}^{k} X_t^2 X_{t+n-k}^2 + k(\overline{X^2})^2 - \overline{X^2}\sum_{t=1}^{k} X_t^2 - \overline{X^2}\sum_{t=1}^{k} X_{t+n-k}^2\right) \Big/ \gamma_{n,X^2}(0).\end{aligned}$$

Here $\overline{X^2} = n^{-1}\sum_{t=1}^{n} X_t^2$ denotes the sample mean of the squared observations. By stationarity, for every fixed t,

$$n^{-1}(1+X_t^2)X_{t+n-k}^2 \xrightarrow{\mathbb{P}} 0, \qquad n \to \infty.$$

Hence it suffices to show that

$$\frac{1+(\overline{X^2})^2+\overline{X^2}}{\gamma_{n,X^2}(0)} \xrightarrow{\mathbb{P}} 0. \tag{8.28}$$

We have

$$\frac{(\overline{X^2})^2}{\gamma_{n,X^2}(0)} = \frac{x_n(\overline{X^2})^2}{x_n\gamma_{n,X^2}(0)}. \tag{8.29}$$

Recall that $\kappa < 4$. According to (8.12), $x_n\gamma_{n,X^2}(0)$ converges in distribution to a positive $\kappa/4$–stable random variable. If $\kappa > 2$, by the ergodicity of the GARCH$(1,1)$ process,

$$(\overline{X^2})^2 \xrightarrow{\text{a.s.}} (\mathbb{E}X_0^2)^2 < \infty.$$

Now, since $\lim_{n\to\infty} x_n = 0$ the sequence in (8.29) converges to zero in probability. If $\kappa \le 2$ then for $0 < \epsilon < \kappa/2$

$$\mathbb{E}[(x_n^{1/2}\overline{X^2})^{\kappa/2-\epsilon}] \le n^{-\kappa/4+\epsilon/2+2\epsilon/\kappa}\,\mathbb{E}|X_0|^{\kappa-2\epsilon}.$$

For small ϵ, the right–hand side converges to zero. Therefore and by Markov's inequality,

$$x_n^{1/2}\overline{X^2} \xrightarrow{\mathbb{P}} 0, \qquad n \to \infty.$$

This shows that (8.29) is asymptotically negligible. The other terms in (8.28) can be treated in a similar way. We omit details. □

Proof of (a). Observe that

$$\begin{aligned} u_{nm}(\boldsymbol{\vartheta}) &= \sum_{|k|<m}\Big(1-\frac{|k|}{m}\Big)\rho_{n,X^2}(k)b_k(\boldsymbol{\vartheta}) + 2\sum_{k=1}^{m-1}\Big(1-\frac{k}{m}\Big)\rho_{n,X^2}(n-k)b_k(\boldsymbol{\vartheta}) \\ &= \sum_{|k|<m}\Big(1-\frac{|k|}{m}\Big)\rho_{n,X^2}(k)b_k(\boldsymbol{\vartheta}) + o_{\mathbb{P}}(1). \end{aligned}$$

The second sum on the first line of the latter display converges to zero in probability uniformly for $\boldsymbol{\vartheta} \in K$, by virtue of Lemma 8.5.3 and since $b_k(\boldsymbol{\vartheta})$ is bounded on K (Lemma 8.5.2). A continuous mapping argument, paired with the weak convergence of the sample ACF, see (8.13), proves $u_{nm} \xrightarrow{\text{d}} v_m$ as $n \to \infty$.

Proof of (b). By Kolmogorov's existence theorem (see Billingsley [12]), we may assume that the sequence of limiting random variables (ρ_h) is defined on a common probability space. Since $\sup_{\boldsymbol{\vartheta}\in K}\sum_k |b_k(\boldsymbol{\vartheta})| < \infty$ and $|\rho_k| \le 1$ a.s., Lebesgue dominated convergence yields

$$v_m(\boldsymbol{\vartheta}) = \sum_{|k|<m} (1-|k|/m)\rho_k b_k(\boldsymbol{\vartheta}) \xrightarrow{\text{a.s.}} \sum_{k\in\mathbb{Z}} \rho_k b_k(\boldsymbol{\vartheta}) = v(\boldsymbol{\vartheta}), \qquad m\to\infty,$$

in $\mathbb{C}(K)$. This proves (b) and concludes the proof of (8.23).

Part (B). Now we turn to the weak limit of the gradient ∇u_n. As a matter of fact, one can follow the lines of the above proof, replacing everywhere the Fourier coefficients b_k by their derivatives ∇b_k and making use of Lemma 8.5.2 for the gradients. Then the same arguments show that

$$\nabla u_n(\boldsymbol{\vartheta}) \xrightarrow{\text{d}} \sum_{k\in\mathbb{Z}} \rho_k \nabla b_k(\boldsymbol{\vartheta}) \qquad \text{in } \mathbb{C}(K,\mathbb{R}^2).$$

It remains to show that one can interchange ∇ and $\sum_{k\in\mathbb{Z}}$ in the limiting process. This follows by an application of Lemma 8.5.2, the fact that $|\rho_k| \le 1$ a.s. and Lebesgue dominated convergence. This proves (8.24), (8.25) and concludes the proof of the proposition. □

Proof of Theorem 8.3.1. As mentioned above, the case $\kappa > 4$ is identical with the one for ARMA processes with iid noise and therefore omitted. Throughout we deal with the case $\kappa < 4$.

The proof is by contradiction. So assume the Whittle estimator *is* consistent, i.e.,

$$\hat{\boldsymbol{\vartheta}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\vartheta}_0. \tag{8.30}$$

By assumption, $\boldsymbol{\vartheta}_0$ is an interior point of C. Therefore we can find a compact set $K\subset C$ such that $\boldsymbol{\vartheta}_0$ is an interior point of K. We conclude from Proposition 8.5.1 that $\nabla u_n \xrightarrow{\text{d}} \nabla v$ in $\mathbb{C}(K,\mathbb{R}^2)$. This, combined with the consistency assumption (8.30) and Corollary 2.4.2, yields that

$$\nabla u_n(\hat{\boldsymbol{\vartheta}}_n) \xrightarrow{\text{d}} \nabla v(\boldsymbol{\vartheta}_0), \qquad n\to\infty.$$

However, $\nabla u_n(\hat{\boldsymbol{\vartheta}}_n) = \mathbf{0}$ as soon as $\hat{\boldsymbol{\vartheta}}_n$ is in the interior of K, and therefore

$$\mathbf{0} = \nabla v(\boldsymbol{\vartheta}_0) = \nabla b_0(\boldsymbol{\vartheta}_0) + 2\sum_{k=1}^{\infty} \rho_k \nabla b_k(\boldsymbol{\vartheta}_0). \tag{8.31}$$

We will show below that (8.31) implies that

$$\nabla b_0(\boldsymbol{\vartheta}_0) = \mathbf{0}. \tag{8.32}$$

On the other hand, $\nabla b_0(\boldsymbol{\vartheta}_0)$ can be calculated directly from

$$b_0(\boldsymbol{\vartheta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1+\varphi_1^2-2\varphi_1\cos(\lambda)}{1+\beta_1^2-2\beta_1\cos(\lambda)}\,\mathrm{d}\lambda = \frac{1+\varphi_1^2+2\varphi_1\beta_1}{1-\beta_1^2},$$

and it is easy to see that $\nabla b_0(\boldsymbol{\vartheta}_0) \neq \mathbf{0}$. This yields the desired contradiction to the consistency assumption (8.30).

Thus it remains to show (8.32). We again proceed by contradiction: assume that $|\nabla b_0(\boldsymbol{\vartheta}_0)| > \delta$ for some $\delta > 0$. By Lemma 8.5.2, $|\nabla b_k(\boldsymbol{\vartheta}_0)|$ decays exponentially fast. Therefore and since $|\rho_k| \leq 1$ a.s., for every $\delta > 0$ one can find $M \geq 1$ such that

$$2\left|\sum_{k=M+1}^{\infty} \rho_k \nabla b_k(\boldsymbol{\vartheta}_0)\right| < \delta/2.$$

Define

$$c = \sum_{k=1}^{M} |\nabla b_k(\boldsymbol{\vartheta}_0)| \quad \text{and} \quad D_M = \left\{\sum_{k=1}^{M} \rho_k \leq \delta/(4c)\right\}.$$

Recall from Remark 8.2.3 that ρ_k is positive with probability 1. Then on D_M,

$$2\left|\sum_{k=1}^{\infty} \rho_k \nabla b_k(\boldsymbol{\vartheta}_0)\right| < \delta,$$

from which we deduce with the triangle inequality that

$$\nabla b_0(\boldsymbol{\vartheta}_0) + 2\sum_{k=1}^{\infty} \rho_k \nabla b_k(\boldsymbol{\vartheta}_0) \neq \mathbf{0} \qquad \text{on } D_M .$$

It remains to show that D_M has positive probability. It was proved in Davis and Mikosch [35] that the limits $\rho_k = V_k/V_0$ are nondegenerate, hence $V_1, \ldots, V_M$ is not a multiple of V_0. The vector $(V_0, \ldots, V_M)$ is jointly $\kappa/4$–stable with all components nondegenerate and positive. Hence $(V_0, \sum_{k=1}^{M} V_k)$ is jointly $\kappa/4$–stable with a Lebesgue density. Therefore, $\mathbb{P}(D_M) > 0$ which finally concludes the proof of Theorem 8.3.1. □

8.6 Proof of Theorem 8.3.2

The proof is similar to the ARMA case with iid innovations; see Brockwell and Davis [29], Section 10.8, for the finite variance and Mikosch et al. [98] for the infinite variance case. As in the latter references, the proof crucially depends on the understanding of the limits of the quadratic forms

$$\sum_j \eta(\lambda_j) I_{n,X^2}(\lambda_j) \tag{8.33}$$

for some appropriate functions η; cf. Proposition 10.8.6 in Brockwell and Davis [29] and Lemma 6.3 in Mikosch et al. [98]. The following result says that, for appropriate functions η, the weak limit of the quadratic forms (8.33) is determined by the weak limits of the sample ACVF of (X_t^2).

Proposition 8.6.1. *Assume the conditions of Theorem 8.3.2 hold. Let $\eta(\lambda)$ be a continuous real–valued 2π–periodic function such that*

(i) $\int_{-\pi}^{\pi} \eta(\lambda)\, g(\lambda; \boldsymbol{\vartheta}_0)\, \mathrm{d}\lambda = 0$,

(ii) *The Fourier coefficients* $f_k = (2\pi)^{-1} \int_{-\pi}^{\pi} \eta(\lambda)\, e^{-ik\lambda} \mathrm{d}\lambda$ *of the function* η *decay geometrically fast, i.e., there exist* $0 < a < 1$ *and* $c > 0$ *such that*

$$|f_k| \le c a^{|k|}, \qquad k \in \mathbb{Z}.$$

Then

$$x_n \left(\frac{1}{n} \sum_j \eta(\lambda_j)\, I_{n,X^2}(\lambda_j) \right) \xrightarrow{\mathrm{d}} \sum_{k \in \mathbb{Z}} f_k V_k, \qquad n \to \infty, \tag{8.34}$$

where $V_k = V_{-k}$ *and* (V_k) *is the distributional limit of the sample ACVF of the squared* GARCH$(1,1)$ *process as specified by* (8.14) *and* (8.16).

The proof will be given at the end of the section.

Proof of Theorem 8.3.2. We proceed analogously to the classical proof as given for Theorem 10.8.2, pp. 390–396, in Brockwell and Davis [29]. A Taylor expansion of $\partial \bar{\sigma}_n^2(\boldsymbol{\vartheta})/\partial \boldsymbol{\vartheta}$ at $\hat{\boldsymbol{\vartheta}}_n$ gives

$$\frac{\partial \bar{\sigma}_n^2(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} = \frac{\partial \bar{\sigma}_n^2(\hat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}} + \frac{\partial^2 \bar{\sigma}_n^2(\tilde{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}^2} (\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) = \frac{\partial^2 \bar{\sigma}_n^2(\tilde{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}^2} (\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0), \tag{8.35}$$

where $|\tilde{\boldsymbol{\vartheta}}_n - \hat{\boldsymbol{\vartheta}}_n| \le |\boldsymbol{\vartheta}_0 - \hat{\boldsymbol{\vartheta}}_n|$. Since $\kappa > 4$, the Whittle estimator is strongly consistent, i.e., $\hat{\boldsymbol{\vartheta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\vartheta}_0$; see Theorem 8.3.1. Therefore $\tilde{\boldsymbol{\vartheta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\vartheta}_0$. The same arguments as for the proof of Proposition 8.5.1 (A) yield that

$$\frac{\partial^2 \bar{\sigma}_n^2(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}^2} \xrightarrow{\text{a.s.}} \frac{\sigma_\nu^2}{2\pi} \int_{-\pi}^{\pi} g(\lambda; \boldsymbol{\vartheta}_0) \frac{\partial^2 (1/g(\lambda; \boldsymbol{\vartheta}))}{\partial \boldsymbol{\vartheta}^2}\, \mathrm{d}\lambda,$$

uniformly on any compact $K \subset C$, where $\sigma_\nu^2 = \mathbb{E}\nu_0^2$ and $(\nu_t) = (X_t^2 - \sigma_t^2)$. The uniformity of convergence and $\tilde{\boldsymbol{\vartheta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\vartheta}_0$ imply that

$$\frac{\partial^2 \bar{\sigma}_n^2(\tilde{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}^2} \xrightarrow{\text{a.s.}} \frac{\sigma_\nu^2}{2\pi} \int_{-\pi}^{\pi} g(\lambda; \boldsymbol{\vartheta}_0)\, \frac{\partial^2 (1/g(\lambda; \boldsymbol{\vartheta}_0))}{\partial \boldsymbol{\vartheta}^2}\, \mathrm{d}\lambda = \mathbf{W}(\boldsymbol{\vartheta}_0), \qquad n \to \infty. \tag{8.36}$$

The last identity is proved on pp. 390–391 in Brockwell and Davis [29].

Since the matrix $\mathbf{W}(\boldsymbol{\vartheta}_0)$ is strictly positive definite with inverse $[\mathbf{W}(\boldsymbol{\vartheta}_0)]^{-1}$, (8.35), (8.36), a continuous mapping and a Cramér–Wold device argument suggest that it suffices to prove the relation

$$\mathbf{c}^T \left(x_n\, \frac{\partial \bar{\sigma}_n^2(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \right) \xrightarrow{\mathrm{d}} \mathbf{c}^T \left(\mathbf{f}_0(\boldsymbol{\vartheta}_0) V_0 + 2 \sum_{k=1}^{\infty} \mathbf{f}_k(\boldsymbol{\vartheta}_0) V_k \right), \tag{8.37}$$

for any $\mathbf{c} \in \mathbb{R}^2$. Observe that

$$\mathbf{c}^T \frac{\partial \bar{\sigma}_n^2(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} = \frac{1}{n} \sum_j \eta(\lambda_j)\, I_{n,X^2}(\lambda_j)\,, \quad \text{where} \quad \eta(\lambda) = \mathbf{c}^T \frac{\partial (1/g(\lambda;\boldsymbol{\vartheta}_0))}{\partial \boldsymbol{\vartheta}}\,.$$

The function η satisfies the conditions of Proposition 8.6.1 as shown on p. 391 of Brockwell and Davis [29]. An application of that proposition proves (8.37) and concludes the proof. □

Proof of Proposition 8.6.1. The main idea is to express the sum on the left–hand side of (8.34) as a linear combination of sample autocovariances of the process (X_t^2) and to apply Theorem 8.2.1 on the asymptotic behavior of the sample ACVF. This idea will be made to work in various steps through a series of lemmas.

Write the left–hand expression of (8.34) as follows:

$$\begin{aligned} & x_n \left(\frac{1}{n} \sum_j \eta(\lambda_j)\, I_{n,X^2}(\lambda_j) \right) \\ &= \frac{x_n}{n} \sum_j \sum_{|h|<n} \eta(\lambda_j)\, \gamma_{n,X^2}(h)\, e^{-ih\lambda_j} \\ &= \frac{x_n}{n} \sum_j \sum_{|h|<n} \eta(\lambda_j)\, (\gamma_{n,X^2}(h) - \gamma_{X^2}(h))\, e^{-ih\lambda_j} \\ &\quad + \frac{x_n}{n} \sum_j \sum_{|h|<n} \eta(\lambda_j)\, \gamma_{X^2}(h)\, e^{-ih\lambda_j} \\ &= I_1 + I_2. \end{aligned} \tag{8.38}$$

Lemma 8.6.2. *Under the assumptions of Proposition 8.6.1, $I_2 \to 0$.*

Proof of Lemma 8.6.2. Note that

$$\sum_j e^{im\lambda_j} = \begin{cases} 0, & \text{if } m \notin n\mathbb{Z}, \\ n, & \text{if } m \in n\mathbb{Z}, \end{cases} \tag{8.39}$$

where the summation is over the Fourier frequencies (4.16). Since $\eta(\lambda) = \sum_{k\in\mathbb{Z}} f_k e^{ik\lambda}$ for all $\lambda \in \mathbb{R}$ and making use of (8.39), we have

$$\begin{aligned}
I_2 &= \frac{x_n}{n} \sum_{k\in\mathbb{Z}} \sum_{|h|<n} \gamma_{X^2}(h) f_k \sum_j e^{i(k-h)\lambda_j} \\
&= x_n \sum_{|h|<n} \gamma_{X^2}(h) \sum_{s\in\mathbb{Z}} f_{h+sn} \\
&= x_n \sum_{|h|<n} \gamma_{X^2}(h) f_h + x_n \sum_{|h|<n} \gamma_{X^2}(h) \sum_{s\in\mathbb{Z}\setminus\{0\}} f_{h+sn} \\
&= I_{21} + I_{22}.
\end{aligned}$$

Observe that by Assumption (i) of Proposition 8.6.1

$$\begin{aligned}
\sum_{h\in\mathbb{Z}} \gamma_{X^2}(h)\, f_h &= \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \gamma_{X^2}(h) \left(\int_{-\pi}^{\pi} \eta(\lambda)\, e^{-ih\lambda} \mathrm{d}\lambda \right) \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} \eta(\lambda) \left(\sum_{h\in\mathbb{Z}} \gamma_{X^2}(h)\, e^{-ih\lambda} \right) \mathrm{d}\lambda \\
&= \sigma_\nu^2 \int_{-\pi}^{\pi} \eta(\lambda)\, g(\lambda; \boldsymbol{\vartheta}_0)\, \mathrm{d}\lambda = 0,
\end{aligned}$$

from which fact it follows that $\sum_{|h|<n} \gamma_{X^2}(h) f_h = -\sum_{|h|\geq n} \gamma_{X^2}(h) f_h$. Recall that both the autocovariances $\gamma_{X^2}(h)$ and the Fourier coefficients decay exponentially fast in $|h|$. Hence

$$\lim_{n\to\infty} I_{21} = -\lim_{n\to\infty} x_n \sum_{|h|\geq n} \gamma_{X^2}(h)\, f_h = 0.$$

The convergence $I_{22} \to 0$ follows from the bounds

$$\left| \sum_{s\in\mathbb{Z}\setminus\{0\}} f_{h+sn} \right| \leq K\, a^{n-|h|}, \qquad |h| < n,$$

for some constant $K > 0$. This concludes the proof. □

We continue to deal with I_1 in (8.38). Again substituting $\eta(\lambda_j)$ by its Fourier series, taking into account (8.39) and setting

$$f_n(h) = \sum_{s\in\mathbb{Z}} f_{h+sn}, \tag{8.40}$$

we obtain

$$I_1 = x_n \sum_{|h|<n} f_n(h)\, (\gamma_{n,X^2}(h) - \gamma_{X^2}(h)). \tag{8.41}$$

For $m \geq 1$, we want to approximate I_1 by

$$I_1(m) = x_n \sum_{|h| \le m} f_n(h)\,(\gamma_{n,X^2}(h) - \gamma_{X^2}(h)).$$

Observe that for every fixed $h \in \mathbb{Z}$,

$$\lim_{n\to\infty} f_n(h) = f_h.$$

Therefore and by virtue of the weak convergence of the sample ACVF, see (8.14) and (8.16), we have

$$I_1(m) = x_n \sum_{|h| \le m} f_n(h)\,(\gamma_{n,X^2}(h) - \gamma_{X^2}(h)) \xrightarrow{\mathrm{d}} \sum_{|h|\le m} f_h\, V_h. \qquad (8.42)$$

Hence by Theorem 2.4.3 it remains to show the following two limit relations:

$$\sum_{|h|\le m} f_h V_h \xrightarrow{\mathrm{d}} \sum_{h\in\mathbb{Z}} f_h V_h, \qquad m \to \infty, \qquad (8.43)$$

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\,(|I_1 - I_1(m)| > \epsilon) = 0 \qquad \text{for all } \epsilon > 0. \qquad (8.44)$$

However, (8.43) follows from (8.44):

Lemma 8.6.3. *Assume* (8.44) *holds. Then the sequence* $\sum_{|h|\le m} f_h V_h$ *has a weak limit as* $m \to \infty$*, which we denote by* $\sum_{h\in\mathbb{Z}} f_h V_h$.

Proof. Since weak convergence is metrized by the Prohorov metric π and the space of distributions on $\mathbb{R}$ is complete (see Billingsley [13], p. 72–73), it is enough to show that the distributions induced by $\sum_{|h|\le m} f_h V_h$ form a Cauchy sequence with respect to the Prohorov metric. We also observe that for any two random variables X, Y the relation $\mathbb{P}(|X - Y| > \epsilon) < \epsilon$ implies $\pi(\mathbb{P}_X, \mathbb{P}_Y) < \epsilon$. Hence

$$\begin{aligned}\lim_{m,k\to\infty} \mathbb{P}\Big(\Big|\sum_{m<|h|\le k} f_h V_h\Big| > \epsilon\Big) &= \lim_{m,k\to\infty} \lim_{n\to\infty} \mathbb{P}(\,|I_1(m) - I_1(k)| > \epsilon)\\ &\le 2 \lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}(\,|I_1(m) - I_1| > \epsilon/2) = 0\end{aligned}$$

for all $\epsilon > 0$ implies that $\sum_{|h|\le m} f_h V_h$ is a Cauchy sequence. □

The proof of (8.44) is quite technical. It is given in the remainder of this section (Proposition 8.6.4) and concludes the proof of Proposition 8.6.1. □

Notice that the coefficients $f_n(h)$ in (8.40) satisfy the following bounds: there exists $K > 0$ such that

$$|f_n(h)| = \Big| f_h + \sum_{s=1}^{\infty} (f_{h+sn} + f_{h-sn}) \Big| \le K\,(a^{|h|} + a^{n-|h|}), \qquad |h| < n.$$

Therefore the desired relation (8.44) follows from the following proposition.

Proposition 8.6.4. *Assume that the conditions of Theorem 8.3.2 hold. Let* $g_n(h)$, $0 \le h < n$, *be numbers satisfying*

$$|g_n(h)| \le K(a^h + a^{n-h}) \tag{8.45}$$

for some constants $K > 0$, $0 < a < 1$. *Then for every* $\epsilon > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(x_n \left| \sum_{h=m+1}^{n} g_n(h) \left(\gamma_{n,X^2}(h) - \gamma_{X^2}(h)\right) \right| > \epsilon \right) = 0. \tag{8.46}$$

Proof of Proposition 8.6.4. Relation (8.46) is equivalent to

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(x_n \left| \sum_{h=m+1}^{n-m} g_n(h) \left(\gamma_{n,X^2}(h) - \gamma_{X^2}(h)\right) \right| > \epsilon \right) = 0 \tag{8.47}$$

for every $\epsilon > 0$. Indeed, an argument similar to the proof of Lemma 8.5.3 shows that for every fixed $h \ge 0$,

$$x_n \left(\gamma_{n,X^2}(n-h) - \gamma_{X^2}(n-h)\right) \xrightarrow{\mathbb{P}} 0.$$

We reduce (8.47) to a simpler problem. Write

$$I_3(m) = x_n \sum_{h=m+1}^{n-m} g_n(h) \left[\left(\gamma_{n,X^2}(h) - \gamma_{X^2}(h)\right) - \frac{1}{n} \sum_{t=1}^{n-h} (X_t^2 X_{t+h}^2 - \mathbb{E}[X_0^2 X_h^2]\,) \right],$$

$$I_4(m) = \frac{x_n}{n} \sum_{h=m+1}^{n-m} g_n(h) \sum_{t=1}^{n-h} (X_t^2 X_{t+h}^2 - \mathbb{E}[X_0^2 X_h^2]\,).$$

Lemma 8.6.5. *The following relation holds:*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\, |I_3(m)| > \epsilon\right) = 0\,. \tag{8.48}$$

Remark 8.6.6. A careful study of the proofs shows that $I_3(m) \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$ for every fixed m provided $4 < \kappa < 8$, whereas we could show only the weaker relation (8.48) for $\kappa \ge 8$. □

Proof of Lemma 8.6.5. We write $I_3(m)$ as follows:

$$
\begin{aligned}
I_3(m) &= -\frac{x_n}{n}\,\overline{X^2}\sum_{h=m+1}^{n-m} g_n(h)\sum_{t=1}^{n-h}(X_t^2-\mathbb{E}X_0^2)\\
&\quad -\frac{x_n}{n}\,\overline{X^2}\sum_{h=m+1}^{n-m} g_n(h)\sum_{t=1}^{n-h}(X_{t+h}^2-\mathbb{E}X_0^2)\\
&\quad +x_n\sum_{h=m+1}^{n-m} g_n(h)\,\frac{n-h}{n}\,[\overline{X^2}-\mathbb{E}X_0^2]^2\\
&\quad -x_n\sum_{h=m+1}^{n-m} g_n(h)\left[1-\frac{n-h}{n}\right]\gamma_{X^2}(h)\\
&= -\overline{X^2}(I_{31}+I_{32})+I_{33}-I_{34}.
\end{aligned}
$$

We have by (8.45),

$$
|I_{34}| \le K\,\frac{x_n}{n}\sum_{h=m+1}^{n-m}(a^h+a^{n-h})\,h\,\gamma_{X^2}(h).
$$

Since the ACVF $\gamma_{X^2}(\cdot)$ decays exponentially fast to zero and $x_n/n \to 0$ we conclude that $I_{34} \to 0$.

The term I_{33} can be treated by observing that the central limit theorem holds;

$$
\sqrt{n}\left(\overline{X^2}-\mathbb{E}X_0^2\right) \xrightarrow{\mathrm{d}} N(0,\sigma^2)
$$

for some positive σ^2. This follows from a standard central limit theorem (see e.g. Ibragimov and Linnik [68]) for strongly mixing sequences with geometric rate; see Boussama [23] for a verification of the latter property in the general GARCH(p,q) case.

Since the terms I_{31} and I_{32} can be treated in the same way we only focus on I_{31}. Its variance is given by

$$
\mathrm{Var}(I_{31}) = \frac{x_n^2}{n^2}\sum_{h=m+1}^{n-m}\sum_{h'=m+1}^{n-m} g_n(h)\,g_n(h')\sum_{t=1}^{n-h}\sum_{t'=1}^{n-h'}\gamma_{X^2}(|t-t'|).
$$

Since $\gamma_{X^2}(h)$ decays exponentially in h, there is a constant $c'>0$ such that

$$
\sum_{t=1}^{n-h}\sum_{t'=1}^{n-h'}\gamma_{X^2}(|t-t'|) \le c'\,n,
$$

and, consequently,

$$\operatorname{Var}(I_{31}) \le \frac{x_n^2}{n^2} \sum_{h=m+1}^{n-m} \sum_{h'=m+1}^{n-m} |g_n(h) g_n(h')| \, (c' \, n) = c' \, \frac{x_n^2}{n} \left(\sum_{h=m+1}^{n-m} |g_n(h)| \right)^2 . \tag{8.49}$$

Recall that (x_n^2/n) is bounded (and converges to zero for $\kappa < 8$). Therefore and in view of condition (8.45) on $g_n(h)$ we conclude that the right–hand side of (8.49) converges to zero by first letting $n \to \infty$ and then $m \to \infty$. This and an application of Markov's inequality conclude the proof. □

By virtue of Lemma 8.6.5 it suffices for (8.47) to show that for every $\epsilon > 0$,

$$\lim_{m \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(|I_4(m)| > \epsilon \right) = 0.$$

We show this by further decomposing $I_4(m)$ into asymptotically negligible pieces.

For ease of notation write

$$\tilde{\gamma}_{n,X^2}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t^2 X_{t+h}^2 - \mathbb{E}[X_0^2 X_h^2]) .$$

Choose a constant $p > 0$ such that

$$\tau = 1 - 4/\min(\kappa, 8) + p \log(a) < 0, \tag{8.50}$$

where we recall that $|g_n(h)| \le K(a^h + a^{n-h})$; see the assumptions of Proposition 8.6.4. Write

$$\begin{aligned} I_4(m) &= x_n \left(\sum_{h=m+1}^{[p\log(n)]} + \sum_{h=[p\log(n)]+1}^{n-[p\log(n)]} + \sum_{h=n-[p\log(n)]+1}^{n-m} \right) g_n(h) \, \tilde{\gamma}_{n,X^2}(h) \\ &= I_{41}(m) + I_{42} + I_{43}(m). \end{aligned}$$

We start by showing that $I_{42} \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. This follows from a simple estimate for the first moment: for some constant $K' > 0$,

$$\begin{aligned} \mathbb{E}|I_{42}| &\le \frac{x_n}{n} \sum_{h=[p\log(n)]}^{n-[p\log(n)]} |g_n(h)| \sum_{t=1}^{n-h} 2\, \mathbb{E}(X_t^2 X_{t+h}^2) \\ &\le 2\,(\mathbb{E}X_0^4)\, x_n \sum_{h=[p\log(n)]}^{n-[p\log(n)]} |g_n(h)| \\ &\le K' \, x_n \, a^{p\log(n)} = K' \, x_n \, n^{p\log(a)} . \end{aligned}$$

The right hand expression is of the order n^τ, where $\tau < 0$ as assumed in (8.50).

Thus it remains to bound $I_{41}(m)$ and $I_{43}(m)$. It suffices to study $I_{41}(m)$ since the other remainder $I_{43}(m)$ can be treated in an analogous way.

In what follows we will use truncation techniques for the summands $X_t^2 X_{t+h}^2 - \mathbb{E}[X_0^2 X_h^2]$. We choose the truncation level a_n in such a way that

$$n\,\mathbb{P}(|X_0| > a_n) \sim n^{-1}.$$

It is then immediate from the tail behavior of $|X_0|$ that one can choose $a_n = (c_0 n)^{1/\kappa}$; see (8.5). Write

$$I_{41}(m) = I_{411}(m) + I_{412}(m)\,,$$

where

$$I_{411}(m) = \frac{x_n}{n} \sum_{h=m+1}^{[p\log(n)]} g_n(h) \sum_{t=1}^{n-h} (X_t^2 X_{t+h}^2 \mathbf{1}_{\{\sigma_t > a_n\}} - \mathbb{E}[X_t^2 X_{t+h}^2 \mathbf{1}_{\{\sigma_t > a_n\}}]\,),$$

$$I_{412}(m) = \frac{x_n}{n} \sum_{h=m+1}^{[p\log(n)]} g_n(h) \sum_{t=1}^{n-h} (X_t^2 X_{t+h}^2 \mathbf{1}_{\{\sigma_t \le a_n\}} - \mathbb{E}[X_t^2 X_{t+h}^2 \mathbf{1}_{\{\sigma_t \le a_n\}}]\,).$$

The treatment of $I_{41i}(m)$ heavily depends on the fact that the volatility process (σ_t^2) satisfies the SRE (8.2), i.e., $\sigma_{t+1}^2 = A_t \sigma_t^2 + B_t$, with $A_t = \alpha_1^\circ Z_t^2 + \beta_1^\circ$ and $B_t = \alpha_0^\circ$. An iteration of this SRE yields the identity

$$\sigma_{t+h}^2 = U_{th} + V_{th}\,\sigma_t^2\,, \quad h \ge 1, \tag{8.51}$$

where

$$U_{th} = \sum_{j=1}^{h-1} A_{t+h-1} \cdots A_{t+j} B_{t+j-1} + B_{t+h-1} \quad \text{and} \quad V_{th} = A_{t+h-1} \cdots A_t.$$

Lemma 8.6.7. *For every $\epsilon > 0$,*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\,(\,|I_{411}(m)| > \epsilon) = 0.$$

Proof of Lemma 8.6.7. By (8.51),

$$X_t^2 X_{t+h}^2 \mathbf{1}_{\{\sigma_t > a_n\}} = \sigma_t^2 \mathbf{1}_{\{\sigma_t > a_n\}} (Z_t^2 Z_{t+h}^2 U_{th}) + \sigma_t^4 \mathbf{1}_{\{\sigma_t > a_n\}} (Z_t^2 Z_{t+h}^2 V_{th}), \tag{8.52}$$

where $Z_t^2 Z_{t+h}^2 U_{th}$ and $Z_t^2 Z_{t+h}^2 V_{th}$ are independent of σ_t^2 for $h \ge 0$. Since $\varphi_1^\circ = \alpha_1^\circ + \beta_1^\circ < 1$ there exists a constant $c > 0$ such that

$$\mathbb{E}[Z_t^2 U_{th} Z_{t+h}^2] = (\mathbb{E}Z_t^2)\,(\mathbb{E}Z_{t+h}^2)\,(\mathbb{E}U_{th}) = \alpha_0^\circ \Big(\sum_{j=1}^{h-1} (\varphi_1^\circ)^{h-j} + 1 \Big) \le c,$$

$$\mathbb{E}[Z_t^2 V_{th} Z_{t+h}^2] = (\mathbb{E}Z_{t+h}^2)\,(\mathbb{E}[V_{th} Z_t^2]) = (\alpha_1^\circ \mathbb{E}Z_0^4 + \beta_1^\circ)\,(\varphi_1^\circ)^{h-1} \le c,$$

for all $h \geq 1$. Taking the expectation in (8.52), we have

$$\mathbb{E}[X_t^2 X_{t+h}^2 \mathbf{1}_{\{\sigma_t > a_n\}}] \leq 2c\,\mathbb{E}[\sigma_0^4 \mathbf{1}_{\{\sigma_0 > a_n\}}]\,, \quad h \geq 1$$

when $a_n \geq 1$. The latter inequality implies that

$$\begin{aligned}\mathbb{E}|I_{411}(m)| &\leq \frac{x_n}{n} \sum_{h=m+1}^{[p\log(n)]} |g_n(h)| \sum_{t=1}^{n-h} 2\,\mathbb{E}[X_t^2 X_{t+h}^2 \mathbf{1}_{\{\sigma_t > a_n\}}] \\ &\leq 4\,c\,x_n\,\mathbb{E}[\sigma_0^4 \mathbf{1}_{\{\sigma_0 > a_n\}}] \sum_{h=m+1}^{[p\log(n)]} |g_n(h)|. \qquad (8.53)\end{aligned}$$

By Karamata's theorem (see e.g. Embrechts et al. [45], Theorem A3.6),

$$x_n\,\mathbb{E}[\sigma_0^4 \mathbf{1}_{\{\sigma_0 > a_n\}}] \sim \text{const.} \qquad (8.54)$$

The Markov inequality together with $\lim_{m\to\infty} \limsup_{n\to\infty} \sum_{h=m+1}^{[p\log(n)]} |g_n(h)| = 0$, (8.53) and (8.54) yield the statement of the lemma. □

We continue with $I_{412}(m)$. Substituting X_t^2 by $\sigma_t^2 Z_t^2$ and X_{t+h}^2 by $(U_{th} + V_{th}\sigma_t^2) Z_{t+h}^2$, we obtain

$$\begin{aligned}&I_{412}(m) \\ &= \frac{x_n}{n} \sum_{h=m+1}^{[p\log(n)]} g_n(h) \sum_{t=1}^{n-h} (\sigma_t^4 \mathbf{1}_{\{\sigma_t \leq a_n\}} Z_t^2 V_{th} Z_{t+h}^2 - \mathbb{E}[\sigma_t^4 \mathbf{1}_{\{\sigma_t \leq a_n\}} Z_t^2 V_{th} Z_{t+h}^2]\,) \\ &+ \frac{x_n}{n} \sum_{h=m+1}^{[p\log(n)]} g_n(h) \sum_{t=1}^{n-h} (\sigma_t^2 \mathbf{1}_{\{\sigma_t \leq a_n\}} Z_t^2 U_{th} Z_{t+h}^2 - \mathbb{E}[\sigma_t^2 \mathbf{1}_{\{\sigma_t \leq a_n\}} Z_t^2 U_{th} Z_{t+h}^2]\,) \\ &= I_{4121}(m) + I_{4122}(m).\end{aligned}$$

The following two lemmas deal with $I_{412i}(m)$, $i = 1, 2$, and conclude the proof of Proposition 8.6.4.

Lemma 8.6.8. *For every* $\epsilon > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\,(\,|I_{4121}(m)| > \epsilon) = 0.$$

Proof of Lemma 8.6.8. Set $S_{th} = Z_t^2 V_{th} Z_{t+h}^2$. We first prove that there is $c > 0$ such that

$$\mathbb{E}S_{th}^2 \leq c \quad \text{for all } t \in \mathbb{Z},\ h \geq 0. \qquad (8.55)$$

From the convexity of the function $g(r) = \mathbb{E}A_0^r$ and $g(\kappa/2) = 1$, see (8.6), we conclude that $g(2) = \mathbb{E}A_0^2 < 1$. Hence

$$\begin{aligned}\mathbb{E}S_{th}^2 &= \mathbb{E}[Z_t^4 A_t^2]\,(\mathbb{E}A_{t+1}^2)\cdots(\mathbb{E}A_{t+h-1}^2)(\mathbb{E}Z_{t+h}^4)\\ &= (\,\mathbb{E}[(\alpha_1^\circ)^2 Z_0^8) + \mathbb{E}(\alpha_1^\circ\beta_1^\circ Z_0^6) + \mathbb{E}[(\beta_1^\circ)^2 Z_0^4]\,)\,(\mathbb{E}A_{t+1}^2)\cdots(\mathbb{E}A_{t+h-1}^2)(\mathbb{E}Z_0^4)\,,\end{aligned}$$

which proves (8.55). Since S_{th} is independent of σ_t^2, we can further decompose

$$I_{4121}(m) = I_{41211}(m) + I_{41212}(m)\,,$$

where

$$I_{41211}(m) = \frac{x_n}{n}\sum_{h=m+1}^{[p\log(n)]} g_n(h)\sum_{t=1}^{n-h}(\sigma_t^4\mathbf{1}_{\{\sigma_t\le a_n\}} - \mathbb{E}[\sigma_t^4\mathbf{1}_{\{\sigma_t\le a_n\}}]\,)\,(\mathbb{E}S_{th}),$$

$$I_{41212}(m) = \frac{x_n}{n}\sum_{h=m+1}^{[p\log(n)]} g_n(h)\sum_{t=1}^{n-h}\sigma_t^4\mathbf{1}_{\{\sigma_t\le a_n\}}(S_{th} - \mathbb{E}S_{th}).$$

One can easily see that the summation in $I_{41211}(m)$ can be extended to $t = 1,\dots,n$ without an impact on the asymptotics. Indeed,

$$\frac{x_n}{n}\sum_{h=m+1}^{[p\log(n)]} g_n(h)\sum_{t=n-h+1}^{n}(\sigma_t^4\mathbf{1}_{\{\sigma_t\le a_n\}} - \mathbb{E}[\sigma_t^4\mathbf{1}_{\{\sigma_t\le a_n\}}]\,)\,(\mathbb{E}S_{th}) \xrightarrow{\mathbb{P}} 0,$$

since the first absolute moment converges to zero. Moreover, we may drop the indicators $\mathbf{1}_{\{\sigma_t\le a_n\}}$ in $I_{41211}(m)$ since for all $\epsilon > 0$,

$$\limsup_{n\to\infty}\mathbb{P}\left(\frac{x_n}{n}\left|\sum_{h=m+1}^{[p\log(n)]} g_n(h)\sum_{t=1}^{n}(\sigma_t^4\mathbf{1}_{\{\sigma_t> a_n\}} - \mathbb{E}[\sigma_t^4\mathbf{1}_{\{\sigma_t> a_n\}}]\,)\,(\mathbb{E}S_{th})\right| > \epsilon\right)$$
$$\to 0,\qquad m\to\infty.$$

This can be shown by computing the first absolute moment of the random variable in the above probability, where one has to account for the asymptotic rate of $\mathbb{E}[\sigma_t^4\mathbf{1}_{\{\sigma_t> a_n\}}]$ in (8.54) and for

$$\lim_{m\to\infty}\limsup_{n\to\infty}\sum_{h=m+1}^{[p\log(n)]}|g_n(h)| \le \lim_{m\to\infty}\limsup_{n\to\infty} K\sum_{h=m+1}^{[p\log(n)]}(a^h + a^{n-h}) = 0. \tag{8.56}$$

Because of these two observations and since $\mathbb{E}S_{th}$ is bounded by $\sqrt{c}$ according to (8.55) and Lyapunov's inequality, it suffices to study the convergence of

$$\tilde{I}_{41211}(m) = \frac{x_n}{n}\sum_{h=m+1}^{[p\log(n)]} g_n(h)\sum_{t=1}^{n}(\sigma_t^4 - \mathbb{E}\sigma_t^4) = x_n\tilde{\gamma}_{n,\sigma^2}(0)\sum_{h=m+1}^{[p\log(n)]} g_n(h). \tag{8.57}$$

It is shown in Section 5.2.2 of Mikosch and Stărică [99] that $x_n\tilde{\gamma}_{n,\sigma^2}(0) \xrightarrow{\text{d}} W$ for some random variable W. This together with (8.56) and a Slutsky argument show that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(|\tilde{I}_{41211}(m)| > \epsilon\right) = 0$$

and therefore

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(|I_{41211}(m)| > \epsilon\right) = 0.$$

It remains to study $I_{41212}(m)$. We will study the second moments:

$$\mathbb{E}[I_{41212}(m)^2] = \frac{x_n^2}{n^2} \sum_{h=m+1}^{[p\log(n)]} \sum_{h'=m+1}^{[p\log(n)]} g_n(h)g_n(h') \sum_{t=1}^{n-h} \sum_{t'=1}^{n-h'} \mathbb{E}[F(t,h,t',h')],$$

where

$$F(t,h,t',h') = \sigma_t^4 \mathbf{1}_{\{\sigma_t\le a_n\}} \sigma_{t'}^4 \mathbf{1}_{\{\sigma_{t'}\le a_n\}} (S_{th} - \mathbb{E}S_{th})(S_{t'h'} - \mathbb{E}S_{t'h'}).$$

Notice that $\mathbb{E}[F(t,h,t',h')] = 0$ whenever $|t-t'| > \min(h,h')$. Indeed, assuming without loss of generality $t' > t$ and $t'-t > h$, $S_{t'h'}$ is independent of $\sigma_{t'}^2, S_{th}, \sigma_t^2$. Then it is straightforward that

$$\mathbb{E}(F(t,h,t',h') \mid \sigma_t^2, \sigma_{t'}^2, S_{th}) = 0 \qquad \text{a.s.}$$

Therefore

$$\mathbb{E}[I_{41212}(m)^2] = \frac{x_n^2}{n^2} \sum_{h=m+1}^{[p\log(n)]} \sum_{h'=m+1}^{[p\log(n)]} g_n(h)g_n(h') \sum_{t=1}^{n-h} \sum_{\substack{t'=1\\ |t'-t|\le\min(h,h')}}^{n-h'} \mathbb{E}[F(t,h,t',h')]\,. \tag{8.58}$$

Note that by Hölder's inequality, independence of σ_t^2 and S_{th}, and (8.55),

$$\begin{aligned} &\mathbb{E}[F(t,h,t',h')] \\ &\le \left(\mathbb{E}[\sigma_t^8 \mathbf{1}_{\{\sigma_t\le a_n\}} (S_{th} - \mathbb{E}S_{th})^2\,]\right)^{1/2} \left(\mathbb{E}[\sigma_{t'}^8 \mathbf{1}_{\{\sigma_{t'}\le a_n\}} (S_{t'h'} - \mathbb{E}S_{t'h'})^2\,]\right)^{1/2} \\ &= \mathbb{E}(\sigma_0^8 \mathbf{1}_{\{\sigma_0\le a_n\}})\,(\mathrm{Var}[S_{th}])^{1/2}\,(\mathrm{Var}[S_{t'h'}])^{1/2} \\ &\le c\,\mathbb{E}(\sigma_0^8 \mathbf{1}_{\{\sigma_0\le a_n\}})\,. \end{aligned} \tag{8.59}$$

In the case $4 < \kappa < 8$, which we will pursue (if $\kappa > 8$ the inequality $\mathbb{E}(\sigma_0^8 \mathbf{1}_{\{\sigma_0\le a_n\}}) \le \mathbb{E}\sigma_0^8 < \infty$ will do), Karamata's theorem gives

$$\mathbb{E}(\sigma_0^8 \mathbf{1}_{\{\sigma_0\le a_n\}}) \sim \text{const} \times a_n^{8-\kappa}, \qquad n\to\infty.$$

This together with (8.58), inequality (8.59) and $\min(h,h') \le \sqrt{hh'}$ leads to

$$\mathbb{E}[I_{41212}(m)^2] \le c' \,\frac{x_n^2}{n^2} \sum_{h=m+1}^{[p\log(n)]} \sum_{h'=m+1}^{[p\log(n)]} |g_n(h)|\,|g_n(h')|\,[2\,n\,\sqrt{hh'}a_n^{8-\kappa}]$$

$$= 2\,c'\,\frac{x_n^2}{n}\,a_n^{8-\kappa}\left(\sum_{h=m+1}^{[p\log(n)]} h^{1/2}|g_n(h)|\right)^2$$

for some $c' > 0$. Note that

$$n^{-1}x_n^2 a_n^{8-\kappa} \sim \text{const}\,.$$

Moreover, since $|g_n(h)| \le K(a^h + a^{n-h})$,

$$\sum_{h=m+1}^{[p\log(n)]} h^{1/2}|g_n(h)| \le K \sum_{h=m+1}^{[p\log(n)]} h^{1/2}a^h + [p\log(n)]^{1/2} \sum_{h=m+1}^{[p\log(n)]} a^{n-h}$$

$$\le K \sum_{h=m+1}^{[p\log(n)]} h^{1/2}a^h + K\,(1-a)^{-1}\,[p\log(n)]^{1/2}\,a^{n-[p\log(n)]}$$

$$\to K \sum_{h=m+1}^{\infty} h^{1/2}a^h \quad \text{as } n\to\infty,$$

$$\to 0 \qquad \text{as } m \to \infty.$$

Therefore

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{E}[I_{41212}(m)^2] = 0,$$

which relation together with Markov's inequality finishes the proof of the lemma. □

It finally remains to show that $I_{4122}(m)$ is negligible.

Lemma 8.6.9. *For every $\epsilon > 0$,*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}(|I_{4122}(m)| > \epsilon) = 0.$$

Proof of Lemma 8.6.9. Let $\tilde S_{th} = Z_t^2 U_{th} Z_{t+h}^2$. Write

$$I_{4122}(m) = I_{41221}(m) + I_{41222}(m),$$

where

$$I_{41221}(m) = \frac{x_n}{n} \sum_{h=m+1}^{[p\log(n)]} g_n(h) \sum_{t=1}^{n-h} (\sigma_t^2 \mathbf{1}_{\{\sigma_t \le a_n\}} - \mathbb{E}[\sigma_t^2 \mathbf{1}_{\{\sigma_t\le a_n\}}])\,\tilde S_{th},$$

$$I_{41222}(m) = \frac{x_n}{n} \sum_{h=m+1}^{[p\log(n)]} g_n(h) \sum_{t=1}^{n-h} \sigma_t^2 \mathbf{1}_{\{\sigma_t\le a_n\}} (\tilde S_{th} - \mathbb{E}\tilde S_{th}\,).$$

Now one can follow the lines of the proof of Lemma 8.6.8 with S_{th} replaced by $\tilde{S}_{th}$. Note that $\mathbb{E}\tilde{S}_{th}$ is also bounded by a constant; set $q = (\mathbb{E}A_0^2)^{1/2}$ and note that $0 < \mathbb{E}A_0 \leq q$ by Lyapunov's inequality. Hence

$$\begin{aligned}
\frac{\mathbb{E}\tilde{S}_{th}^2}{2(\alpha_0^\circ)^2(\mathbb{E}Z_0^4)^2} &= \frac{\mathbb{E}U_{th}^2}{2(\alpha_0^\circ)^2} \\
&\leq \frac{1}{2(\alpha_0^\circ)^2}\mathbb{E}\left[2\left(\sum_{j=1}^{h-1} A_{t+h-1}\cdots A_{t+j}B_{t+j-1}\right)^2 + 2B_{t+h-1}^2\right] \\
&= 1 + \mathbb{E}\left[\sum_{j=1}^{h-1}\sum_{j'=1}^{h-1} A_{t+h-1}\cdots A_{t+j}\cdot A_{t+h-1}\cdots A_{t+j'}\right] \\
&= 1 + \sum_{j=1}^{h-1}\sum_{j'=1}^{h-1} (\mathbb{E}A_0^2)^{h-\max(j,j')}(\mathbb{E}A_0)^{|j-j'|} \\
&\leq 1 + \sum_{j=1}^{h-1}\sum_{j'=1}^{h-1} q^{2h-2\max(j,j')}q^{|j-j'|} \\
&= 1 + \sum_{j=1}^{h-1}\sum_{j'=1}^{h-1} q^{2h-j-j'} \\
&= 1 + \left(\sum_{j=1}^{h-1} q^{h-j}\right)^2 .
\end{aligned}$$

Secondly the term corresponding to (8.57) in Lemma 8.6.8 has to be treated by the central limit theorem, see also Lemma 8.6.5. □

Remark 8.6.10. As a matter of fact, the only place in the proof of Proposition 8.6.4, where we made use of the assumption $\mathbb{E}Z_0^8 < \infty$, was the proof of Lemma 8.6.8. We conjecture that this assumption can be replaced by $\mathbb{E}|Z_0|^{\kappa+\delta} < \infty$ for some positive δ.

□

References

1. H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)*. Akadémiai Kiadó, Budapest, 1973.
2. D. W. Andrews. Generic uniform convergence. *Econometric Theory*, 8:241–257, 1992.
3. M. Babillot, P. Bougerol, and L. Elie. The random difference equation $x_n = a_n x_{n-1} + b_n$ in the critical case. *Ann. Probab.*, 25:478–493, 1997.
4. R. Baillie and T. Bollerslev. The message in daily exchange rates: a conditional variance tale. *J. Bus. Econom. Statist.*, 7:297–305, 1989.
5. B. Basrak, R. A. Davis, and T. Mikosch. Regular variation of GARCH processes. *Stochastic Process. Appl.*, 99:95–115, 2002.
6. I. Berkes and L. Horváth. The rate of consistency of the quasi–maximum likelihood estimator. *Statist. Probab. Lett.*, 61:133–143, 2003.
7. I. Berkes and L. Horváth. The efficiency of the estimators of the parameters in GARCH processes. *Ann. Statist.*, 32:633–655, 2004.
8. I. Berkes, L. Horváth, and P. Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, 9:201–228, 2003.
9. P. J. Bickel and P. Bühlmann. What is a linear process ? *Proc. Natl. Acad. Sci. USA*, 93:12128–12131, 1996.
10. P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore, 1993.
11. P. Billingsley. *Statistical Inference for Markov Processes.* University of Chicago Press, 1961.
12. P. Billingsley. *Probabiliy and Measure.* Wiley, New York, third edition, 1995.
13. P. Billingsley. *Convergence of Probability Measures.* Wiley, New York, second edition, 1999.
14. N. H. Bingham, C. Goldie, and J. Teugels. *Regular Variation*, volume 27 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, 1987.
15. G. D. Birkhoff. Proof of the ergodic theorem. *Proc. Natl. Acad. Sci. USA*, 17:656–660, 1931.
16. F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. Polit. Economy*, 81:637–654, 1973.

17. S. Bochner. Integration von Funktionen, deren Werte die Elemente eines Vectorraumes sind. *Fund. Math.*, 20:262–276, 1933.
18. T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31:307–327, 1986.
19. T. Bollerslev. A conditionally heteroskedastic time series model for security prices and rates of return data. *Rev. Econom. Statist.*, 69:542–547, 1987.
20. P. Bougerol. Kalman filtering with random coefficients and contractions. *SIAM J. Control Optim.*, 31:942–959, 1993.
21. P. Bougerol and N. Picard. Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics*, 52:115–127, 1992.
22. P. Bougerol and N. Picard. Strict stationarity of generalized autoregressive processes. *Ann. Probab.*, 20:1714–1730, 1992.
23. F. Boussama. Ergodicité, mélange et estimation dans les modèles GARCH. PhD Thesis, Université 7, Paris, 1998.
24. R. C. Bradley. Basic properties of strong mixing conditions. In E. Eberlein and M. Taqqu (Eds.): *Dependence in Probability and Statistics.* Birkhäuser, Boston, 1986.
25. A. Brandt. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. in Appl. Probab.*, 18:211–220, 1986.
26. J. Breidt, R. A. Davis, and A. Trindade. Least absolute deviation estimation for all–pass time series models. *Ann. Statist.*, 29:919–946, 2001.
27. L. Breiman. On some limit theorems similar to the arc–sin law. *Teor. Verojatnost. i Primenen.*, 10:351–360, 1965.
28. P. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting.* Springer, New York, fifth edition, 1996.
29. P. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer, New York, fifth edition, 1996.
30. P. Bühlmann and A. J. McNeil. An algorithm for nonparametric GARCH modelling. *Comput. Statist. Data Anal.*, 40:665–683, 2002.
31. M. Carrasco and X. Chen. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18:17–39, 2002.
32. H. Cramér. *Mathematical Methods of Statistics.* Princeton University Press, 1946.
33. M. Dacorogna, R. Gençay, U. Müller, R. Olsen, and O. Pictet. *An Introduction to High–Frequency Finance.* Academic Press, San Diego, 2001.
34. R. A. Davis and W. Dunsmuir. Maximum likelihood estimation for MA(1) processes with a root on or near the unit circle. *Econometric Theory*, 12:1–29, 1996.
35. R. A. Davis and T. Mikosch. The sample autocorrelations of heavy–tailed processes with applications to ARCH. *Ann. Statist.*, 26:2049–2080, 1998.
36. R. A. Davis and S. Resnick. Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.*, 13:179–195, 1985.
37. R. A. Davis and S. Resnick. Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.*, 14:533–558, 1986.
38. P. Diaconis and D. Freedman. Iterated random functions. *SIAM Rev.*, 41:45–76, 1999.
39. Z. Ding, C. W. J. Granger, and R. Engle. A long memory property of stock market returns and a new model. *J. Empirical Finance*, 1:83–106, 1993.

40. P. Doukhan. *Mixing: Properties and Examples*, volume 85 of *Lecture Notes in Statistics*. Springer, New York, 1994.
41. F. Drost and G. González-Rivera. Efficiency comparison of maximum–likelihood–based estimators in GARCH models. *J. Econometrics*, 93:93–111, 1999.
42. F. Drost and C. A. J. Klaassen. Efficient estimation in semiparametric GARCH models. *J. Econometrics*, 81:193–221, 1997.
43. R. M. Dudley. *Real Analysis and Probability.* Cambridge University Press, Cambridge, second edition, 2002.
44. K. Dzhaparidze. *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series.* Springer, New York, 1986.
45. P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin, first edition, 1997.
46. R. Engle. Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1007, 1982.
47. R. Engle and G. González-Rivera. Semiparametric ARCH models. *J. Bus. Econom. Statist.*, 9:345–359, 1991.
48. E. Fama. The behavior of stock market prices. *J. Bus.*, 38:34–105, 1965.
49. P. D. Feigin and R. L. Tweedie. Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.*, 6:1–14, 1985.
50. W. Feller. *An Introduction to Probability Theory and its Applications, Vol. I.* Wiley, New York, third edition, 1968.
51. T. S. Ferguson. *A Course in Large Sample Theory.* Chapman and Hall, London, 1996.
52. W. A. Fuller. *Introduction to Statistical Time Series.* Wiley, New York, second edition, 1996.
53. E. Ghysels, A. Harvey, and E. Renault. Stochastic volatility. In *Handbook of Statistics* 14. North-Holland, Amsterdam, 1996.
54. I. Gijbels, A. Pope, and M. Wand. Understanding exponential smoothing via kernel regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 61:39–50, 1999.
55. L. Giraitis, P. Kokoszka, and R. Leipus. Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory*, 16:3–22, 2000.
56. L. Giraitis and P. M. Robinson. Whittle estimation of ARCH models. *Econometric Theory*, 17:608–631, 2001.
57. L. Glosten, R. Jagannathan, and D. Runkle. On the relation between expected value and the volatility of the nominal excess return on stocks. *J. Finance*, 48:1779–1801, 1993.
58. C. Goldie. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.*, 1:126–166, 1991.
59. C. Gouriéroux and A. Monfort. *Statistics and Econometric Models.* Cambridge University Press, 1996.
60. C. Gouriéroux, A. Monfort, and A. Trognon. Pseudomaximum likelihood methods: theory. *Econometrica*, 52:681–700, 1984.
61. C. W. J. Granger and A. Andersen. On the invertibility of time series models. *Stochastic Process. Appl.*, 8:87–92, 1978.
62. P. Hall and C. C. Heyde. *Martingale Limit Theory and its Applications.* Academic Press, New York, 1980.
63. P. Hall and Q. Yao. Inference in ARCH and GARCH models with heavy–tailed errors. *Econometrica*, 71:285–317, 2003.

64. E. J. Hannan. The asymptotic theory of linear time–series models. *J. Appl. Probab.*, 10:130–145, 1973.
65. E. J. Hannan and M. Deistler. *The Statistical Theory of Linear Systems.* Wiley, New York, 1988.
66. L. Horváth and F. Liese. L_p-estimators in ARCH models. *J. Statist. Plann. Inference*, 119:277–309, 2004.
67. P. J. Huber. *Robust Statistics.* Wiley, New York, 1981.
68. I. A. Ibragimov and Y. V. Linnik. *Independent and Stationary Sequences of Random Variables.* Wolters–Noordhoff, Groningen, 1971.
69. Insightful. *S+FinMetrics.* See `http://www.insightful.com`.
70. T. Jeantheau. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, 14:70–86, 1998.
71. O. Kallenberg. *Foundations of Modern Probability.* Springer, New York, second edition, 2002.
72. H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Math.*, 131:207–248, 1973.
73. H. Kesten and F. Spitzer. Convergence in distribution of products of random matrices. *Z. Wahrsch. Verw. Gebiete*, 67:363–386, 1984.
74. J. F. C. Kingman. Subadditive ergodic theory. *Ann. Probab.*, 1:883–909, 1973.
75. L. Klimko and P. I. Nelsen. On conditional least squares estimation for stochastic processes. *Ann. Statist.*, 6:629–642, 1978.
76. P. Kokoszka and M. Taqqu. Parameter estimation for infinite variance fractional ARIMA. *Ann. Statist.*, 24:1880–1913, 1996.
77. U. Krengel. *Ergodic Theorems.* de Gruyter, Berlin, 1985.
78. A. Kufner, O. John, and S. Fučík. *Function Spaces.* Noordhoff, Leyden, 1977.
79. S. Kullback and R. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22:79–86, 1951.
80. S. Lang. *Analysis* I/II. Addison–Wesley, Boston, 1969.
81. L. Le Cam. On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. California Publ. Statist.*, 1:277–329, 1953.
82. M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes.* Springer, Berlin, 1983.
83. M. Ledoux and M. Talagrand. *Probability in Banach Spaces.* Springer, Berlin, 1991.
84. S. Lee and B. Hansen. Asymptotic theory for the GARCH(1,1) quasi–maximum likelihood estimator. *Econometric Theory*, 10:29–52, 1994.
85. E. Lehmann. *Theory of Point Estimation.* Wiley, New York, 1983.
86. G. Letac. A contraction principle for certain Markov chains and its applications. In J. E. Cohen, H. Kesten, and C. M. Newman (Eds.): *Random Matrices and their Applications (Brunswick, Maine, 1984).* Amer. Math. Soc., Providence RI, 1986.
87. A. Lindner and K. Meyer. Extremal behaviour of finite EGARCH processes. Preprint Munich University of Technology, 2002.
88. S. Ling and M. McAleer. On adaptive estimation in nonstationary ARMA models with GARCH errors. *Ann. Statist.*, 31:642–674, 2003.
89. R. Lumsdaine. Finite–sample properties of the maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models: a Monte Carlo investigation. *J. Bus. Econom. Statist.*, 13:1–10, 1995.

90. R. Lumsdaine. Consistency and asymptotic normality of the quasi–maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica*, 64:575–596, 1996.
91. J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley, Chichester, 1988.
92. B. Mandelbrot. The variation of certain speculative prices. *J. Bus.*, 36:394–419, 1963.
93. H. Markowitz. *Portfolio Selection: Efficient Diversification of Investments.* Wiley, New York, 1959.
94. A. J. McNeil. *EVIS: Software for Extreme Values in S–PLUS.* Downloadable at `http://www.math.ethz.ch/~mcneil/software.html`.
95. A. J. McNeil and R. Frey. Estimation of tail–related risk measures for heteroscedastic financial time series: an extreme value approach. *J. Empirical Finance*, 7:271–300, 2000.
96. S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Springer, London, 1993.
97. T. Mikosch. Modeling dependence and tails of financial time series. In B. Finkenstädt and H. Rootzén (Eds.): *Extreme Value Theory and Applications.* Chapman and Hall, London, 2003.
98. T. Mikosch, T. Gadrich, C. Klüppelberg, and R. Adler. Parameter estimation for ARMA models with infinite variance innovations. *Ann. Statist.*, 23:305–326, 1995.
99. T. Mikosch and C. Stărică. Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Ann. Statist.*, 28:1427–1451, 2000.
100. T. Mikosch and C. Stărică. Change of structure in financial time series and the GARCH model. *Revstat*, 2:16–41, 2004.
101. T. Mikosch and C. Stărică. Non-stationarities in financial time series, the long-range dependence and the IGARCH effects. *Rev. Econ. Statist.*, 86:378–390, 2004.
102. T. Mikosch and D. Straumann. Whittle estimation in a heavy–tailed GARCH(1, 1) model. *Stochastic Process. Appl.*, 100:187–222, 2002.
103. T. Mikosch and D. Straumann. Stable limits of martingale transforms with application to parameter estimation in conditionally heteroscedastic time series models. Preprint University of Copenhagen, 2003.
104. A. Mokkadem. Propriétés de mélange des processus autorégressifs polynomiaux. *Ann. Inst. H. Poincaré Probab. Statist.*, 26:219–260, 1990.
105. D. Nelson. Stationarity and persistence in the GARCH(1, 1) model. *Econometric Theory*, 6:318–334, 1990.
106. D. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59:347–370, 1991.
107. J. Neveu. *Discrete Parameter Martingales.* North–Holland, Amsterdam, 1975.
108. W. Newey and D. Steigerwald. Asymptotic bias for quasi–maximum–likelihood estimators in conditional heteroskedasticity models. *Econometrica*, 65:587–599, 1997.
109. S. Ould Ahmedou Voffal. L'estimateur des moindres carrés dans les modèles ARCH. *C. R. Acad. Sci. Paris Sér. I Math.*, 322:971–974, 1996.
110. K. R. Parthasarathy. *Probability Measures on Metric Spaces.* Academic Press, New York, 1967.
111. T. Parthasarathy. *Selection Theorems and their Applications*, volume 263 of *Lecture Notes in Mathematics.* Springer, Berlin, 1972.

112. J. Pfanzagl. On the measurability and consistency of minimum contrast estimates. *Metrika*, 14:249–272, 1969.
113. D. T. Pham. The mixing property of bilinear and generalised random coefficient autoregressive models. *Stochastic Process. Appl.*, 23:291–300, 1986.
114. M. B. Priestley. *Spectral Analysis and Time Series.* Academic Press, London, ninth edition, 1996.
115. R. Ranga Rao. Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.*, 33:659–680, 1962.
116. S. I. Resnick. Point processes, regular variation and weak convergence. *Adv. in Appl. Probab.*, 18:66–138, 1986.
117. S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes.* Springer, New York, 1987.
118. E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants.* Springer, Berlin, 2000.
119. RiskMetrics Group. *The RiskMetrics© Technical Document.* Downloadable at `http://www.riskmetrics.com/rmcovv.html`.
120. W. Rudin. *Real and Complex Analysis.* McGraw-Hill, New York, third edition, 1987.
121. G. Samorodnitsky and M. Taqqu. *Stable Non–Gaussian Random Processes. Stochastic Models with Infinite Variance.* Chapman and Hall, London, 1994.
122. P. A. Samuelson. Mathematics of speculative prices. *SIAM Rev.*, 15:1–42, 1973.
123. N. Shephard. Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. Hinkley, and O. E. Barndorff–Nielsen (Eds.) *Likelihood, Time Series with Econometric and Other Applications.* Chapman and Hall, London, 1996.
124. W. Stout. *Almost Sure Convergence.* Academic Press, New York, 1974.
125. D. Straumann and T. Mikosch. Quasi–maximum–likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. Preprint University of Copenhagen, 2003.
126. S. Taylor. *Modelling Financial Time Series.* Wiley, Chichester, 1986.
127. D. Tjøstheim. Estimation in nonlinear time series models. *Stochastic Process. Appl.*, 21:251–273, 1986.
128. D. Tjøstheim. Nonlinear time series and Markov chains. *Adv. in Appl. Probab.*, 22:587–611, 1990.
129. A. W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, 1998.
130. J. von Neumann. On rings of operators. Reduction theory. *Ann. of Math.*, 50:401–485, 1949.
131. A. Wald. Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.*, 20:595–601, 1949.
132. H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–25, 1982.
133. P. Whittle. Estimation and information in stationary time series. *Ark. Mat.*, 2:423–434, 1953.
134. K. Yosida. *Functional Analysis.* Springer, New York, 1965.
135. J.-M. Zakoïan. Threshold heteroscedastic models. *J. Econom. Dynam. Control*, 18:931–955, 1994.
136. E. Zivot and J. Wang. *Modeling Financial Time Series with S-PLUS.* Springer, New York, 2002.
137. G. Zumbach. The pitfalls in fitting GARCH(1, 1) processes. In C. Dunis (Ed.): *Advances in Quantitative Asset Management*, Studies in Computational Finance. Kluwer, Boston, 2000.

Author Index

Index

Lecture Notes in Statistics

For information about Volumes 1 to 128, please contact Springer-Verlag

129: Wolfgang Härdle, Gerard Kerkyacharian, Dominique Picard, and Alexander Tsybakov, Wavelets, Approximation, and Statistical Applications. xvi, 265 pp., 1998.

130: Bo-Cheng Wei, Exponential Family Nonlinear Models. ix, 240 pp., 1998.

131: Joel L. Horowitz, Semiparametric Methods in Econometrics. ix, 204 pp., 1998.

132: Douglas Nychka, Walter W. Piegorsch, and Lawrence H. Cox (Editors), Case Studies in Environmental Statistics. viii, 200 pp., 1998.

133: Dipak Dey, Peter Müller, and Debajyoti Sinha (Editors), Practical Nonparametric and Semiparametric Bayesian Statistics. xv, 408 pp., 1998.

134: Yu. A. Kutoyants, Statistical Inference For Spatial Poisson Processes. vii, 284 pp., 1998.

135: Christian P. Robert, Discretization and MCMC Convergence Assessment. x, 192 pp., 1998.

136: Gregory C. Reinsel, Raja P. Velu, Multivariate Reduced-Rank Regression. xiii, 272 pp., 1998.

137: V. Seshadri, The Inverse Gaussian Distribution: Statistical Theory and Applications. xii, 360 pp., 1998.

138: Peter Hellekalek and Gerhard Larcher (Editors), Random and Quasi-Random Point Sets. xi, 352 pp., 1998.

139: Roger B. Nelsen, An Introduction to Copulas. xi, 232 pp., 1999.

140: Constantine Gatsonis, Robert E. Kass, Bradley Carlin, Alicia Carriquiry, Andrew Gelman, Isabella Verdinelli, and Mike West (Editors), Case Studies in Bayesian Statistics, Volume IV. xvi, 456 pp., 1999.

141: Peter Müller and Brani Vidakovic (Editors), Bayesian Inference in Wavelet Based Models. xiii, 394 pp., 1999.

142: György Terdik, Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis: A Frequency Domain Approach. xi, 258 pp., 1999.

143: Russell Barton, Graphical Methods for the Design of Experiments. x, 208 pp., 1999.

144: L. Mark Berliner, Douglas Nychka, and Timothy Hoar (Editors), Case Studies in Statistics and the Atmospheric Sciences. x, 208 pp., 2000.

145: James H. Matis and Thomas R. Kiffe, Stochastic Population Models. viii, 220 pp., 2000.

146: Wim Schoutens, Stochastic Processes and Orthogonal Polynomials. xiv, 163 pp., 2000.

147: Jürgen Franke, Wolfgang Härdle, and Gerhard Stahl, Measuring Risk in Complex Stochastic Systems. xvi, 272 pp., 2000.

148: S.E. Ahmed and Nancy Reid, Empirical Bayes and Likelihood Inference. x, 200 pp., 2000.

149: D. Bosq, Linear Processes in Function Spaces: Theory and Applications. xv, 296 pp., 2000.

150: Tadeusz Calinski and Sanpei Kageyama, Block Designs: A Randomization Approach, Volume I: Analysis. ix, 313 pp., 2000.

151: Håkan Andersson and Tom Britton, Stochastic Epidemic Models and Their Statistical Analysis. ix, 152 pp., 2000.

152: David Ríos Insua and Fabrizio Ruggeri, Robust Bayesian Analysis. xiii, 435 pp., 2000.

153: Parimal Mukhopadhyay, Topics in Survey Sampling. x, 303 pp., 2000.

154: Regina Kaiser and Agustín Maravall, Measuring Business Cycles in Economic Time Series. vi, 190 pp., 2000.

155: Leon Willenborg and Ton de Waal, Elements of Statistical Disclosure Control. xvii, 289 pp., 2000.

156: Gordon Willmot and X. Sheldon Lin, Lundberg Approximations for Compound Distributions with Insurance Applications. xi, 272 pp., 2000.

157: Anne Boomsma, Marijtje A.J. van Duijn, and Tom A.B. Snijders (Editors), Essays on Item Response Theory. xv, 448 pp., 2000.

158: Dominique Ladiray and Benoît Quenneville, Seasonal Adjustment with the X-11 Method. xxii, 220 pp., 2001.

159: Marc Moore (Editor), Spatial Statistics: Methodological Aspects and Some Applications. xvi, 282 pp., 2001.

160: Tomasz Rychlik, Projecting Statistical Functionals. viii, 184 pp., 2001.

161: Maarten Jansen, Noise Reduction by Wavelet Thresholding. xxii, 224 pp., 2001.

162: Constantine Gatsonis, Bradley Carlin, Alicia Carriquiry, Andrew Gelman, Robert E. Kass Isabella Verdinelli, and Mike West (Editors), Case Studies in Bayesian Statistics, Volume V. xiv, 448 pp., 2001.

163: Erkki P. Liski, Nripes K. Mandal, Kirti R. Shah, and Bikas K. Sinha, Topics in Optimal Design. xii, 164 pp., 2002.

164: Peter Goos, The Optimal Design of Blocked and Split-Plot Experiments. xiv, 244 pp., 2002.

165: Karl Mosler, Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach. xii, 280 pp., 2002.

166: Hira L. Koul, Weighted Empirical Processes in Dynamic Nonlinear Models, Second Edition. xiii, 425 pp., 2002.

167: Constantine Gatsonis, Alicia Carriquiry, Andrew Gelman, David Higdon, Robert E. Kass, Donna Pauler, and Isabella Verdinelli (Editors), Case Studies in Bayesian Statistics, Volume VI. xiv, 376 pp., 2002.

168: Susanne Rässler, Statistical Matching: A Frequentist Theory, Practical Applications and Alternative Bayesian Approaches. xviii, 238 pp., 2002.

169: Yu. I. Ingster and Irina A. Suslina, Nonparametric Goodness-of-Fit Testing Under Gaussian Models. xiv, 453 pp., 2003.

170: Tadeusz Calinski and Sanpei Kageyama, Block Designs: A Randomization Approach, Volume II: Design. xii, 351 pp., 2003.

171: D.D. Denison, M.H. Hansen, C.C. Holmes, B. Mallick, B. Yu (Editors), Nonlinear Estimation and Classification. x, 474 pp., 2002.

172: Sneh Gulati, William J. Padgett, Parametric and Nonparametric Inference from Record-Breaking Data. ix, 112 pp., 2002.

173: Jesper Møller (Editor), Spatial Statistics and Computational Methods. xi, 214 pp., 2002.

174: Yasuko Chikuse, Statistics on Special Manifolds. xi, 418 pp., 2002.

175: Jürgen Gross, Linear Regression. xiv, 394 pp., 2003.

176: Zehua Chen, Zhidong Bai, Bimal K. Sinha, Ranked Set Sampling: Theory and Applications. xii, 224 pp., 2003

177: Caitlin Buck and Andrew Millard (Editors), Tools for Constructing Chronologies: Crossing Disciplinary Boundaries, xvi, 263 pp., 2004

178: Gauri Sankar Datta and Rahul Mukerjee, Probability Matching Priors: Higher Order Asymptotics, x, 144 pp., 2004

179: D.Y. Lin and P.J. Heagerty , Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data, vii, 336 pp., 2004

180: Yanhong Wu, Inference for Change-Point and Post-Change Means After a CUSUM Test, xiv, 176 pp., 2004

181: Daniel Straumann, Estimation in Conditionally Heteroscedastic Time Series Models, xvi, 228 pp., 2005