VILNIUS UNIVERSITY

ANDRIUS ŠKARNULIS

QUADRATIC ARCH MODELS WITH LONG MEMORY AND QML ESTIMATION

Doctoral Dissertation
Physical Sciences, Mathematics (01 P)

Vilnius, 2017

The dissertation was written in 2012–2016 at Vilnius University.

Scientific Supervisor – Prof. Dr. Habil. Donatas Surgailis (Vilnius University, Physical Sciences, Mathematics – 01 P).

Scientific Adviser – Prof. Dr. Habil. Remigijus Leipus (Vilnius University, Physical Sciences, Mathematics – 01 P).

VILNIAUS UNIVERSITETAS

ANDRIUS ŠKARNULIS

KVADRATINIAI ILGOSIOS ATMINTIES ARCH MODELIAI IR PARAMETRŲ VERTINIMAS KVAZIDIDŽIAUSIO TIKĖTINUMO METODU

Daktaro disertacija
Fiziniai mokslai, matematika (01 P)

Vilnius, 2017

Disertacija rengta 2012–2016 metais Vilniaus universitete.

Mokslinis vadovas – prof. habil. dr. Donatas Surgailis (Vilniaus universitetas, fiziniai mokslai, matematika – 01 P).

Mokslinis konsultantas – prof. habil. dr. Remigijus Leipus (Vilniaus universitetas, fiziniai mokslai, matematika – 01 P).

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my scientific supervisor Prof. Donatas Surgailis for his inspiration, patient mentoring and all the support and help during the years of my PhD studies. Without his active involvement, immense knowledge and guidance it would be hard to imagine the results of this dissertation as they are. Also a very special thanks goes to Prof. Liudas Giraitis for his helpful advice, encouragement, thought-provocing discussions and the opportunity to participate in joint work on interesting topics. It has been a great privilege and honor to learn from some of the brightest figures in the field of mathematics. I would also like to thank my loving parents, brother and my fiancée Ugnė for their support and understanding through all these years.

Andrius Škarnulis
Vilnius
April 10, 2017

Table of Contents

Notations and Abbreviations

The following notations and abbreviations are used throughout the dissertation:

$\mathbb{Z}$	set of integers
$\mathbb{N}$	set of positive integers
$\mathbb{R} := (-\infty, \infty)$	set of real numbers
$\mathbb{R}^+$	set of positive real numbers
$\mathbb{C}$	set of complex numbers
$:=$	"by definition"
L	backshift operator ($LX_k = X_{k-1}$)
$[s]$	integer part of s
$\mathrm{Var}(X_k)$	variance of a random variable X_k
$a_n \sim b_n$	if $a_n/b_n \to 1$, $n \to \infty$
C	generic constant
$\|\cdot\|$	the norm (of function or sequence)
d	long memory parameter
Θ	set of model parameters

$\widehat{\theta}_n$	estimator of $\theta \in \Theta$ given *infinite* past (using the unobserved process)
$\widetilde{\theta}_n$	estimator of $\theta \in \Theta$ given *finite* past (using the observed process only)
$\Pi := [-\pi, \pi]$	
$i.i.d.$	independent identically distributed
$\{B(t)\}$	standard Brownian motion
$\{B_{d+1/2}(t)\}$	fractional Brownian motion
$\zeta_k, \xi_k, \varepsilon_k$	usually random "noise" of the process
$\to_{D[0,1]}$	weak convergence in the Skorohod space $D[0,1]$
$\stackrel{P}{\to}$	convergence in probability
$\stackrel{a.s.}{\to}$	almost sure convergence
$\stackrel{d}{\to}$	convergence in distribution
$\Gamma(\cdot)$	Gamma function
$B(\cdot, \cdot)$	Beta function
$E(X_k)$	mean of a random variable X_k
$\mathrm{cov}(\cdot, \cdot), \gamma(\cdot)$	covariance function of a random process
$\mathrm{corr}(\cdot, \cdot), \rho(\cdot)$	correlation function of a random process
∂_x	derivative with respect to variable x
$a.e.$	almost everywhere
$r.h.s.$	right-hand side
$l.h.s.$	left-hand side
$w.r.t.$	with respect to
$r.v.$	random variable
$RMSE$	root mean square error
$QMLE$	quasi-maximum likelihood estimation (estimator)

$ARCH\ process$	Autoregressive Conditionally Heteroscedastic process
$LARCH\ process$	Linear ARCH process
$GQARCH\ process$	Generalized Quadratic ARCH process

All (in)equalities involving random variables in this dissertation are supposed to hold almost surely.

Chapter 1

Introduction

Nowadays the importance of data and information that derives from it is undeniable and grows rapidly. As time passes, scientists, corporations, government institutions and others have the possibility to deal with longer and longer time series of data. It is important to have adequate tools that would allow us to analyze, model and forecast data retrieved from long time series. In view of the importance of economics and financial markets for today's well-being, scientists developed a variety of statistical methods that help better understand the dynamics and behavior of various financial and economic indicators, as well as financial markets and economics in general. It is commonly known that the dynamics of financial markets is dynamical in itself, i.e. the volatility changes over time. However, this important feature – referred to as "conditional heteroscedasticity" in the context of time series – was often dismissed or ignored in many statistical settings and modeling. Robert F. Engle won the 2003 Sveriges Riksbank prize in Economic Sciences in Memory of Alfred Nobel "for methods of analyzing economic time series with time-varying volatility (ARCH)". The parametric ARCH model, introduced by Engle [24] in 1982,

$$r_t = \zeta_t \sigma_t, \quad \sigma_t^2 = \omega + \sum_{j=1}^{q} b_j r_{t-j}^2, \quad t \in \mathbb{Z},$$

where $\{\zeta_t, t \in \mathbb{Z}\}$ is an i.i.d. noise with zero mean and unit variance, was extended in 1986 by Bollerslev [10] to the well-known GARCH(p, q) process $r_t = \zeta_t \sigma_t$ with conditional variance

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} b_j r_{t-j}^2 + \sum_{j=1}^{p} a_j \sigma_{t-j}^2.$$

The concept of autoregressive conditionally heteroscedastic models is widepsread in theory on time series and practical applications.

Later on, parametric models were extended to include information from infinite past, for example, in 1991 Robinson [62] introduced the ARCH(∞) process whose conditional variance has the form

$$\sigma_t^2 = \omega + \sum_{j=1}^{\infty} b_j r_{t-j}^2.$$

In the case of the GARCH(p, q) model, autocovariances decay exponentially fast, ARCH(∞) may have autocovariances decaying to zero at a slower rate, however, in these traditional settings, the stationary solution with finite variance has summable autocovariances: $\sum_{k \in \mathbb{Z}} \gamma(k) < \infty$ (except, as we prove in this dissertation, in the case of ARCH(∞) with $\omega = 0$ and $\sum_{j=1}^{\infty} b_j = 1$), or covariance short memory, which is a major drawback of similar models in light of the well-known phenomenon of long memory in, for example, squared financial returns. As there is a need for new models and there are still important unresolved problems in terms of models that gained popularity in practical applications (e.g. FIGARCH), the topic of long memory conditionally heteroscedastic time series is important, relevant and interesting in itself. The present dissertation focuses on specific ARCH-type models with long memory, as well as the long memory "generating mechanism", from which long memory originates in the ARCH setting. In terms of practical applications, questions related to parameter estimation for long memory models are also under the scope of this dissertation.

The main aims set and problems raised in this dissertation are as follows.

Finding conditions for the existence of a finite variance stationary solution with long memory of FIGARCH, ARCH(∞) and integrated AR(∞) processes (Chapter 3). The main goal is to find the necessary and sufficient conditions for the existence of a stationary solution of the integrated ARCH(∞) process, in particular, the so-called FIGARCH equation, proposed by Baillie, Bollerslev and Mikkelsen [3] in 1996 to capture the long memory effect in volatility. There is much discussion on controversies surrounding the FIGARCH equation. In 1996 Ding and Granger [20] introduced the LM(d)-ARCH model, whose important particular case is the FIGARCH equation. They argued that a stationary solution of the LM(d)-ARCH equation with the finite fourth moment *has a long memory*, however, the existence of such a solution was never shown. By finding the above-mentioned conditions for the existence of a stationary solution, we solve the long standing Ding and Granger conjecture. We also aim to explore the relation between the stationary solutions of ARCH(∞) (as well as FIGARCH) and integrated AR(∞) processes – the stationary solution of the former process is constructed in terms of the solution of the latter.

Questions surrounding the IAR(∞) model are of independent interest. The class of stationary IAR(∞) processes with long memory is vast and, as our simulations show, its special case of IAR(p, d, q) models might be reasonably considered as a new class of long memory models which provides more flexibility to model long memory processes changing their autocovariances on low lags without an effect on the long-term behavior.

Exploring the parametric quasi-maximum likelihood estimation for a new generalized quadratic ARCH (GQARCH) process (Chapter 4). The Quadratic ARCH (QARCH) process with long memory, introduced by Doukhan *et al.* [22], and generalized in Chapter 5 of this dissertation (see also Grublytė and Škarnulis [40]), extends the QARCH model of Sentana [66] and the Linear ARCH (LARCH) model of Robinson [62] to the strictly positive conditional variance. The GQARCH and LARCH models have similar long memory and leverage properties and can both be used to model financial data with these properties. The main disadvantage of the LARCH model in comparison to the GQARCH model is the fact that volatility in the case of LARCH may assume negative values and is not separated from below by positive constant. The standard quasi-maximum likelihood (QML) approach to the estimation of LARCH parameters is inconsistent. We aim to investigate the QML estimation for the 5-parametric GQARCH model, whose parametric form of moving average coefficients is the same as that by Beran and Schützner [5] for the LARCH model. Our main goal is to prove the consistency and asymptotic normality of the corresponding estimates, including long memory parameter $0 < d < 1/2$. Also, a simulation study to evaluate the finite sample performance of the QML estimation for GQARCH model is performed.

Investigating the existence and properties of a stationary solution of the gen-

eralized nonlinear model for long memory conditional heteroscedasticity (Chapter 5). As a parametric ARCH(q) model of Engle [24] was generalized to GARCH(p, q) by Bollerslev [10], we aim to extend the ARCH-type model discussed by Doukhan, Grublytė and Surgailis [22] to the model where conditional variance satisfies an AR(1) equation $\sigma_t^2 = Q^2(a + \sum_{j=1}^{\infty} b_j r_{t-j}) + \gamma\sigma_{t-1}^2$ with a Lipschitz function $Q(x)$.

The novelty of the results in this dissertation:

- conditions for the existence of the stationary finite variance solution of integrated ARCH(∞) and FIGARCH processes with long memory.
- the final answer to the long standing conjecture of Ding and Granger [20] about the existence of a stationary solution of the Long Memory ARCH (as well as FIGARCH) model with long memory and the finite fourth moment.
- introduction and investigation of a new class of long memory integrated AR(p, d, q) processes, whose autocovariance can be modeled easily at low lags without a significant effect on the long memory behavior, this being a major advantage over classical ARFIMA models.
- proof of consistency and asymptotic normality of the QML estimator for the Generalized Quadratic ARCH process, empirical evaluation of the finite sample performance of the QML estimation for the GQARCH model.
- conditions for the existence of a stationary finite variance solution of the generalized nonlinear model for long memory conditional heteroscedasticity, its long memory and leverage properties.

Publications and conferences. The following three papers cover the main results presented in this dissertation:

- L. Giraitis, D. Surgailis and A. Škarnulis. Stationary integrated ARCH(∞) and AR(∞) processes with finite variance. *Submitted*, 2017.
- I. Grublytė, D. Surgailis and A. Škarnulis. QMLE for quadratic ARCH model with long memory. *Journal of Time Series Analysis.* 2016. doi: 10.1111/jtsa.12227.
- I. Grublytė and A. Škarnulis. A nonlinear model for long memory conditional heteroscedasticity. *Statistics.* 51:123–140, 2017.

The main results of this dissertation were also presented at the following conferences:

- 8th International Conference of the ERCIM WG on Computational and Methodological Statistics/9th International Conference on Computational and Financial Econometrics, University of London, 12–14 December, 2015. Title of presentation: *Quasi-MLE for quadratic ARCH model with long memory*.
- NBER-NSF Time Series Conference, Vienna University of Economics and Business, 25–26 September, 2015. Title of presentation: *Integrated AR and ARCH processes and the FIGARCH model: origins of long memory*.
- 11th International Vilnius Conference on Probability Theory and Mathematical Statistics, Vilnius University, 29 June–1 July, 2014. Title of presentation: *An autoregressive conditional duration model and the FIGARCH equation*.

Chapter 2

Background

In Section 2.1 of this chapter we provide some basic definitions and propositions, which will be used throughout the dissertation. Long memory, as an object and important thematic line of this dissertation, will be briefly described in Section 2.2. We touch upon the main principles of the parameter estimation for time series models in Section 2.3.

2.1 Definitions and preliminaries

Since the main interest of this dissertation lies in conditionally heteroscedastic models, first we recall that a time series $\{X_k, k \in \mathbb{Z}\}$ is called *conditionally homoscedastic* if its conditional variance

$$\sigma_k^2 = \mathrm{Var}\,(X_k \mid X_{k-1}, X_{k-2}, ...) = C, \quad k \in \mathbb{Z},$$

is constant, while in terms of *conditionally heteroscedastic* time series, its conditional variance is a random process (in general). In this dissertation, the term "stationary process" is mostly used by means of *covariance* *stationarity.*

Definition 2.1. *Random process $\{X_k, k \in \mathbb{Z}\}$ is called covariance stationary if $\mathrm{E}X_k = C$ and $\mathrm{E}X_k^2 < \infty$ are constant for all $k \in \mathbb{Z}$, and the covariance function*

$$\mathrm{cov}(X_k, X_{k+j}) = \mathrm{cov}(X_0, X_j),$$

is constant in k, for all $k, j \in \mathbb{Z}$.

As it will be stated in Section 2.2 of this chapter, the main instruments we use (in this dissertation) to characterize the long memory property of time series are the covariance and spectral density functions. Proposition 2.1 below describes the relation between the covariance function, spectral distribution function and spectral density. Suppose that function $F : [-\pi, \pi] \to [0, \infty)$ is right-continuous, nondecreasing, bounded and $F(-\pi) = 0$.

Proposition 2.1. *Function $\gamma(k), k \in \mathbb{Z}$, is a covariance function of some stationary process if and only if*

$$\gamma(k) = \int_{-\pi}^{\pi} e^{iks} dF(s),$$

with some (unique) function F, which is called a spectral distribution function. If $F(s) = \int_{\pi}^{s} f(\nu)d\nu$, then f is called a spectral density function.

The concept of a transfer function is used to prove important results of this dissertation (Chapter 3).

Definition 2.2. *Suppose that process $\{X_k, k \in \mathbb{Z}\}$ can be written as*

$$X_k = \sum_{j=0}^{\infty} a_j Z_{k-j}, \quad k \in \mathbb{Z},$$

where $\{Z_k, k \in \mathbb{Z}\}$ *is a stationary process. Then the Fourier transform* $A(x) := \sum_{k=0}^{\infty} e^{-ixk} a_k, x \in \Pi$, *is called the transfer function.*

Being the main object of this dissertation, the ARCH(∞) process is defined as follows.

Definition 2.3. *A nonnegative random process* $\{\tau_k,\ k \in \mathbb{Z}\}$ *is said to satisfy an ARCH(*∞*) equation if there exists a sequence of nonnegative i.i.d. random variables* $\{\varepsilon_k, k \in \mathbb{Z}\}$ *with unit mean* $\mathrm{E}\varepsilon_0 = 1$, *a nonnegative number* $\omega \geq 0$ *and a deterministic sequence* $b_j \geq 0$, $j = 1, 2, \ldots$, *such that*

$$\tau_k \ = \ \varepsilon_k \left(\omega + \sum_{j=1}^{\infty} b_j \tau_{k-j} \right), \qquad k \in \mathbb{Z}. \tag{2.1}$$

In this dissertation, we assume that ARCH-type processes $\{\tau_k, k \in \mathbb{Z}\}$ (or r_k, x_k, depending on notation in a particular context) are *causal*, i.e. for any k, τ_k can be represented as a measurable function $f(\varepsilon_k, \varepsilon_{k-1}, \ldots)$ of the present and past values of innovations $\varepsilon_s, s \leq k$. For example, if stationarity and causality are not required, equation (2.1) can have infinitely many solutions (see, e.g., Leipus and Kazakevičius [51]).

Definition 2.4. *Let* $\{\varepsilon_k, k \in \mathbb{Z}\}$ *be a process of uncorrelated random variables with zero mean and variance* σ_ε^2. *Then a random process* $\{X_k, k \in \mathbb{Z}\}$ *is said to be causal with respect to* $\{\varepsilon_k, k \in \mathbb{Z}\}$ *if* $X_k = f(\varepsilon_k, \varepsilon_{k-1}, ...)$ *for every* $k \in \mathbb{Z}$, *where* f *is a measurable function such that* X_k *is a properly defined random variable.*

An important statistical concept, which will be assigned to many processes considered in this dissertation, is ergodicity. To put in a simple manner, this feature allows estimating the characteristics of a random process, having only one sufficiently long realization of the process, without the need of using multiple independent samples. One often refers to ergodicity for the mean, in which case:

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to \mathrm{E}(X_0) = \mu, \quad n \to \infty.$$

A random process $\{X_k, k \in \mathbb{Z}\}$ is said to be ergodic for the second moment if

$$\frac{1}{n-j}\sum_{k=j+1}^{n} (X_k - \mu)(X_{k-j} - \mu) \xrightarrow{P} \gamma(j), \text{ for all } j,$$

where $\gamma(j) = \operatorname{cov}(X_k, X_{k-j})$ (see, e.g., Hamilton [43]). One can define an ergodic process in a wider sense (see, e.g., Andersen and Moore [2]): a random process $\{X_k, k \in \mathbb{Z}\}$ is ergodic if for any suitable function $f(\cdot)$ the following limit exists almost surely:

$$\mathrm{E}\left[f(X_0)\right] = \lim_{N\to\infty} \frac{1}{2N+1}\sum_{k=-N}^{N} f(X_k).$$

Now we provide a more formal definition of a stationary ergodic time series (see Lindner [54]).

Definition 2.5. *Let $\{X_k, k \in \mathbb{Z}\}$ be a stationary time series of random variables X_k in $\mathbb{R}$. Then $\{X_k, k \in \mathbb{Z}\}$ can be seen as a random element in $\mathbb{R}^{\mathbb{Z}}$, equipped with its Borel-σ-algebra $B(\mathbb{R}^{\mathbb{Z}})$. Let the backshift operator $\Phi : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ be given by $\Phi(\{z_i, i \in \mathbb{Z}\}) = \{z_{i-1}, i \in \mathbb{Z}\}$. Then the time series $\{X_k, k \in \mathbb{Z}\}$ is called ergodic if, for $\Lambda \in B(\mathbb{R}^{\mathbb{Z}})$, $\Phi(\Lambda) = \Lambda$ implies $P(\{X_k, k \in \mathbb{Z}\} \in \Lambda) \in \{0, 1\}$.*

The following proposition about the ergodicity of a random process is a simplified version of Theorem 3.5.8 by Stout [67] and states that a measurable function of an ergodic process forms again an ergodic process.

Proposition 2.2. *Suppose* $\{X_k, k \in \mathbb{Z}\}$ *is an ergodic sequence (e.g. i.i.d. random variables) and* $f : \mathbb{R}^\infty \to \mathbb{R}$ *is a measurable function. Then the sequence* $\{Y_k, k \in \mathbb{Z}\}$*, where*

$$Y_k = f(X_k, X_{k-1}, ...),$$

is an ergodic process.

Since convergence in mean-square is almost without exception used in the definitions of stationary solutions of many models in this dissertation, we give a short definition for this mode of convergence.

Definition 2.6. *We say that the sequence* $\{X_k, k \in \mathbb{Z}\}$ *of square integrable random variables converges in mean-square if there exists a square integrable random variable* X *such that*

$$\lim_{k\to\infty} \mathrm{E}\left[(X_k - X)^2\right] = 0.$$

In this dissertation, phrases "converges in L^2" and "converges in mean-square" are used interchangeably. Similarly to Definition 2.6, one could define the convergence in L^p.

As discussed in Chapter 3 of this dissertation, the stationary solution of the ARCH(∞) process can be constructed in terms of the discrete time infinite Volterra series. For example, we show that the stationary solution of ARCH(∞) process can be written in the form of causal Volterra series:

$$\begin{aligned} Y_k &= \mu + \\ &+ \mu\sigma \left(\sum_{m=1}^{\infty} \sum_{-\infty < s_m < \cdots < s_1 \le k} g_{k-s_1} h_{s_1-s_2} \cdots h_{s_{m-1}-s_m} \zeta_{s_1} \cdots \zeta_{s_m} \right), \quad k \in \mathbb{Z}, \end{aligned} \tag{2.2}$$

with standardized i.i.d. innovations $\{\zeta_k, k \in \mathbb{Z}\}$. In order to correctly define the convergence of the discrete time infinite Volterra series (e.g.

having the form (2.2)), we first remind a few facts and definitions related to the summability in Banach spaces (see, e.g., Hunter and Nachtergaele [45]). First, recall that a normed linear space is a metric space with respect to metric d derived from its norm, where $d(x, y) = \|x - y\|$. A Banach space is a normed linear space that is a complete metric space with respect to the metric derived from its norm.

Definition 2.7. *Let $\{x_i, i \in I\}$ be an indexed set in a Banach space E, where I is a countable index set. For each finite subset J of I, we define the partial sum S_J by*

$$S_J = \sum_{i \in J} x_i.$$

We say that $x \in E$ is a sum of an indexed set $\{x_i, i \in I\}$ if for every $\epsilon > 0$ there is a finite subset J^ϵ of I such that $\|S_J - x\| < \epsilon$ for all finite subsets J of I that contain J^ϵ.

Definition 2.8. *If $x \in E$ is the sum of an indexed set $\{x_i, i \in I\}$ (in the sense of Definition 2.7), then we write $x = \sum_{i \in I} x_i$, and the set $\{x_i, i \in I\}$ is called summable.*

The fact that the set in a Banach space is summable ensures many useful features, for example, the possibility to change the summation order, etc.

Recall that the set U of vectors in a Hilbert space H is orthonormal if it is orthogonal, i.e. for every $x, y \in U$ we have $\langle x, y \rangle = 0$, and $\|x\| = 1$ for all $x \in U$. Let I be the same countable index set as in Definitions 2.7 and 2.8 above, $\{e_i, i \in I\}$ – some orthonormal set in a Hilbert space H (in particular, in an L^2 space), and $\{c_i, i \in I\}$ the set of real numbers. It is known that the square summability of c_i, i.e. $\sum_i c_i^2 < \infty$, guarantees that the set $\{c_i e_i, i \in I\}$ is summable in space H (in particular, in space L^2).

The last fact is especially useful when thinking about the convergence (in L^2) of the discrete time infinite Volterra series (e.g. having the form (2.2)). Indeed, since $\{\zeta_k, k \in \mathbb{Z}\}$ in (2.2) is the sequence of independent and identically distributed standardized random variables, then for each $k \in \mathbb{Z}$ the set $\{\zeta_{s_1} \cdots \zeta_{s_m}, m \geq 1, s_m < ... < s_1 \leq k \in \mathbb{Z}\}$ is orthonormal in L^2. In this case, the convergence of Volterra series mainly depends on its coefficients – if they are square summable (which is ensured by our assumptions, see Chapter 3), the Volterra series converges in L^2. We also note that if variables X_k are nonnegative, then from the summability of $\sum_{k\in\mathbb{Z}} X_k$ in the L^2 space follows the almost sure convergence of this series.

Next we define two classes of processes – bilinear and linear ARCH – which will act as important models in the three main chapters of this dissertation, especially Chapter 3, for the construction of a stationary solution for the IARCH(∞) process. ARCH-type bilinear models where considered by, for example, Giraitis and Surgailis [30].

Definition 2.9. *We say that the discrete stationary process $\{X_k, k \in \mathbb{Z}\}$ satisfies the bilinear equation (or is a bilinear process) if*

$$X_k = \zeta_k \left(a + \sum_{j=1}^{\infty} a_j X_{k-j} \right) + b + \sum_{j=1}^{\infty} b_j X_{k-j}, \tag{2.3}$$

where $\{\zeta_k, k \in \mathbb{Z}\}$ is a sequence of i.i.d. random variables with $\mathrm{E}\zeta_k = 0$ and $\mathrm{Var}(\zeta_k) = 1$, $a, a_j, b, b_j, j \geq 1$, are real coefficients. In the case of $b = b_j \equiv 0$, (2.3) is the Linear ARCH (LARCH) model introduced by Robinson [62].

We define the Brownian motion and the fractional Brownian motion, following Giraitis *et al.* [36].

Definition 2.10. *A Brownian motion is a Gaussian process* $\{B(t), t \in \mathbb{R}^+\}$ *with* $B(0) = 0, \mathrm{E}B(t) \equiv 0$ *and a covariance function* $\gamma_B(s,t) = \mathrm{E}(B_s B_t) = \min(s,t)$*.*

Definition 2.11. *Let* $0 < H < 1$ *be any number. Then a Gaussian process* $\{B_H(t), t \in \mathbb{R}^+\}$ *with* $B_H(0) = 0, \mathrm{E}B_H(t) \equiv 0$ *and a covariance function* $\gamma_H(s,t) := \frac{1}{2}\left\{|s|^{2H} + |t|^{2H} - |s-t|^{2H}\right\}$ *is called a fractional Brownian motion with a Hurst parameter* $0 < H < 1$*.*

2.2 Long memory

The goal of this section is to provide the intuition behind *long memory* as an object of research on time series. From first glance, one might say that the concept of long memory in papers considering long memory stochastic processes mainly refers to slowly decaying autocovariances of the process, i.e. covariance between distant members of the process disappears slowly with an increasing lag between them. Although the so-called second-order properties of the process indeed prevailed in definitions and description of long memory, in general, however, there is a wide diversity of definitions of long memory as such.

An often-used starting point in enclosing the rise of the phenomenon and concept of long memory in scientific literature are the observations by Hurst ([46], [47]). As a hydrologist, he investigated the characteristics of water flow in the river Nile, which is known, among others, for its specific long-term behavior regarding long periods of dryness and yearly returning floods. Hurst considered the possibility to regularize the flow of the Nile. Without elaborating further, we just mention that data was analyzed using the so-called rescaled adjusted range or the R/S-statistic of the form:

$$\frac{\max\limits_{0\le i\le k} X_{t,k} - \min\limits_{0\le i\le k} X_{t,k}}{\left(k^{-1}\sum_{i=t+1}^{t+k}(X_i - \bar{X}_{t,k})^2\right)^{1/2}},$$

where $X_{t,k} = X_{t+i} - X_t - \frac{i}{k}(X_{t+k} - X_t)$. The main message is as follows. For the stationary ergodic sequence $\{X_1, X_2, ...\}$, the statistic R/S grows as the square root of the sample size, that is, $n^{1/2}$. However, in terms of the data on the Nile, considered by Hurst, the R/S empirically grew as $n^{0.74}$. This finding is referred to as the Hurst effect or the Hurst phenomenon. Yet the question is what stochastic process could be used to explain and model the Hurst effect. For example, the attempt to relax the condition of finite variance was unsuccessful (Moran [59]). Mandelbrot with co-authors ([55], [56]), using the Fractional Gaussian Noise, succeeded in modeling the Hurst effect, the main reason behind that being the introduction of long memory in the setting. It is also interesting that from here comes the name of the Hurst parameter H in the fractional Brownian motion (see Definition 2.11).

Popularity of the second-order properties (asymptotic behavior of covariances, spectral density, etc.) in definitions of long memory was underpinned by historical and practical reasons (mainly conceptual simplicity and rather easy estimation from the data). One firstly thinks about slow decay or nonsummability of autocovariances when exploring the long memory property in terms of second-order properties of processes. However, this case is mainly restricted to covariance stationary stochastic processes.

Next we provide several definitions of the long memory property. Similar ones can be found in Giraitis *et al.* [36], Beran [4], Cox [17], Giraitis and Surgailis [30], Giraitis *et al.* [35].

Definition 2.12. *A covariance stationary process* $\{X_k, k \in \mathbb{Z}\}$ *with an autocovariance function* $\gamma(k) = \mathrm{cov}(X_0, X_k)$ *is said to have:*
Covariance long memory if

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| = \infty;$$

Covariance short memory if

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty \textit{ and } \sum_{k\in\mathbb{Z}} \gamma(k) > 0;$$

Negative memory if

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty \textit{ and } \sum_{k\in\mathbb{Z}} \gamma(k) = 0.$$

To take one step further, the above definition can be specified in terms of the asymptotic behavior or the decay rate of the covariance function. For this we need a definition of a slowly varying function.

Definition 2.13. *A function* $L : [0, \infty) \to \mathbb{R}$ *is said to be slowly varying at infinity, if L is positive on* $[a, \infty)$ *(and positive or negative on* $[0, a)$*), for some* $a > 0$*, and*

$$\lim_{x\to\infty} \frac{L(sx)}{L(x)} = 1, \quad \forall s > 0.$$

Definition 2.14. *A function* $f(x), x \geq 0$*, is said to be a regularly varying function with index* $\delta \in \mathbb{R}$*, if* f *is positive on* $[a, \infty)$*, for some* $a > 0$*, and* $\forall s > 0$

$$\lim_{x\to\infty} \frac{f(sx)}{f(x)} = s^{\delta}, \quad \forall s > 0.$$

A regularly varying function f can be written in the form of $f(x) = x^{\delta} L(x)$ for some slowly varying function L. For the majority of models considered in this dissertation, covariance and spectral density functions can be expressed as regularly varying functions of the form $\gamma(k) = |k|^{-1+2d} L(|k|)$ and $f(s) = |s|^{-2d} L(1/|s|), 0 < d < 1/2$.

Definition 2.15. *A stationary process $\{X_k, k \in \mathbb{Z}\}$ has long memory, if its covariance function $\gamma(k) = \mathrm{cov}(X_0, X_k)$ decays hyperbolically to zero:*

$$\gamma(k) = |k|^{2d-1} L(|k|), \quad \forall k \geq 1, \tag{2.4}$$

with a memory parameter $0 < d < 1/2$ and a slowly varying function L.

Condition (2.4) is often specified in a simpler form $\gamma(k) \sim c_{\gamma} |k|^{2d-1}$. Hyperbolically decaying autocovariances are nonsummable (see, e.g., Giraitis *et al.* [36]).

The above definitions are often treated as a long memory characterization in the time domain. Memory definitions in the frequency domain are based on features of spectral density.

Definition 2.16. *Suppose that a stationary process $\{X_k, k \in \mathbb{Z}\}$ has a spectral density function f, which is bounded on $[\varepsilon, \pi]$ for any $\varepsilon > 0$, and satisfies*

$$f(x) = |x|^{-2d} L(1/|v|), \quad v \in \Pi, \tag{2.5}$$

for some slowly varying function L. The process $\{X_k, k \in \mathbb{Z}\}$ is said to have negative memory, or short memory, or long memory, if accordingly $-1/2 < d < 0$, or $d = 0$, or $0 < d < 1/2$.

Condition (2.5) is often simplified to $f(x) \sim c_f |x|^{-2d}, x \to 0$, with some constant $c_f > 0$.

Papers that investigate long memory processes, often alongside consideration of the decay rate and summability of autocovariances, also investigate the convergence of the partial sums process. This is yet another way to define long memory.

Definition 2.17. *We say that a strictly stationary process $\{X_k, k \in \mathbb{Z}\}$ has distributional long memory if its normalized partial sums process*

$$\left\{ A_n^{-1} \sum_{k=1}^{[ns]} (X_k - B_n) : s \in [0,1] \right\}$$

converges, in the sense of weak convergence of the finite dimensional distributions, as $n \to \infty$, to a random process $\{Z(s)\}_{s\in[0,1]}$ with dependent increments. Here $A_n \to \infty, n \to \infty$, and B_n are some constants .

There are many other types of definitions, however, we will not consider them any further. In this dissertation, by long memory we mean the covariance long memory, unless stated otherwise.

2.3 Estimation

The field of statistical procedures and methods to estimate parameters of time series models can be a brigde between theory and practical application. In this section, we briefly discuss and review the main methods used to estimate parameters of conditionally heteroscedastic time series models, not necessarily those with long memory. A variety of different ways was introduced to estimate the time series models. The first two concepts that should be mentioned with reference to this topic are the Least Squares (LS) method and the Quasi-maximum likelihood (QML) method. The former is often called the simplest method to estimate parametric ARCH(q) models, while the latter is particularly relevant for GARCH(p, q). For example, for the strictly stationary GARCH process, the QML estimators are consistent and asymptotically normal with no moment assumption on the observed process, using some mild regularity conditions instead. This is particularly important from a practical point of view as for many financial time series the requirement of the finite fourth or even higher moments is questionable. To provide the main idea behind LS and QML estimation, we use the examples of parametric ARCH and GARCH models. Then we will move on to discussing the case of infinite order models such as ARCH(∞).

The basic idea behind the parameter estimation of the time series model is as follows. Having a finite data set of size n, $\{r_1, ..., r_n\}$, we assume that these observations come from a random process of a specific form which often (but not always) depends on a finite number of parameters. In this section, we denote the true (unknown) values of these parameters with $\theta_0 = (\theta_{01}, ..., \theta_{0p}), p < \infty$. The main goal is to get the "best" estimates $\widehat{\theta}_n$ of θ_0 from the data that we have. Here, the subscript n indicates that we calculate the estimator using the available data sample of size n. In most cases, different methods can be applied. Independently of what we choose, two concepts (features), which are inevitably found in the statistical inference and estimation literature, are a) consistency and b) asymptotic normality of estimators. The estimator is called consistent if $\widehat{\theta}_n \to \theta_0$ in probability as $n \to \infty$. Strong consistency means that the above-mentioned convergence holds almost surely. Most often by asymptotic normality we mean that the difference between the consistent estimator $\widehat{\theta}_n$ and true parameters θ_0 converges in distribution to the normal distribution:

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, V).$$

When $\theta_0 \in \mathbb{R}^p$, then V is a $p \times p$ matrix.

The parametric ARCH(q) model is often used to explain how the Least Squares (LS) method works (see, e.g., Francq and Zakoian [27], Francq and Zakoian [28]). One of the reasons behind this is that, for ARCH(q), the LS estimation provides estimators in the explicit form. Let's consider the process

$$r_k = \sigma_k \zeta_k, \quad \sigma_k^2 = \omega_0 + \sum_{j=1}^{q} a_{0j} r_{k-j}^2, \quad k \in \mathbb{Z}, \tag{2.6}$$

with $\omega_0 > 0$, $a_{0i} \geq 0$, $i = 1, ..., q$, and $\{\zeta_k, k \in \mathbb{Z}\}$ an i.i.d. sequence with zero mean and unit variance. The vector of true parameters is $\theta_0 = (\omega_0, a_{01}, ..., a_{0q})^T$ (T denotes the transposed vector). The LS estimation procedure for ARCH(q) is performed rewriting (2.6) as an AR(q) equation for r_k^2:

$$r_k^2 = \omega_0 + \sum_{j=1}^{q} a_{0j} r_{k-j}^2 + u_k,$$

with $u_k = r_k^2 - \sigma_k^2 = (\zeta_k^2 - 1)\sigma_k^2$. As usual in terms of estimation, we try to estimate the model parameters from a finite sample of observed values $(r_1, ..., r_n)$, with the initial set of observations being $(r_0, ..., r_{1-q})$, all of which can be, for example, zero-valued. The LS estimator is given by

$$\widehat{\theta}_n = (\hat{\omega}, \hat{a}_1, ..., \hat{a}_q) = (X^T X)^{-1} X^T Y,$$

where

$$Y = X\theta_0 + U,$$

with

$$X^T = \begin{pmatrix} Z_{n-1}^T \\ \vdots \\ Z_0^T \end{pmatrix}, \quad Y = \begin{pmatrix} r_n^2 \\ \vdots \\ r_1^2 \end{pmatrix}, \quad U = \begin{pmatrix} u_n \\ \vdots \\ u_1 \end{pmatrix},$$

and vectors $Z_{k-1}^T = (1, r_{k-1}^2, ..., r_{k-q}^2)$.

If $\{r_k, k \in \mathbb{Z}\}$ is a nonanticipative strictly stationary solution of (2.6), $\omega_0 > 0$ and $\mathrm{E}r_k^2 < \infty$, then the LS estimator of $\sigma_0^2 = \mathrm{Var}(u_k)$ is

$$\widehat{\sigma}^2 = \frac{1}{n-q-1} \sum_{t=1}^{n} \left(r_t^2 - \widehat{\omega} - \sum_{j=1}^{q} \widehat{a}_j r_{t-j}^2 \right)^2.$$

Strong consistency, that is,

$$\widehat{\theta}_n \overset{a.s.}{\to} \theta_0, \quad \widehat{\sigma}_n^2 \overset{a.s.}{\to} \sigma_0^2,$$

can be achieved under $\mathrm{E}r_k^4 < \infty$ and $\mathrm{P}(\zeta_k^2 = 1) \neq 1$, while for asymptotic normality the finiteness of the eight moment is needed, $\mathrm{E}r_k^8 < \infty$ (see, e.g., Bose and Mukherjee [11]), then

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \overset{d}{\to} N(0, (\mathrm{E}\zeta_k^4 - 1)A^{-1}BA^{-1}),$$

where $A = \mathrm{E}(Z_q Z_q^T)$ and $B = \mathrm{E}(\sigma_{q+1}^4 Z_q Z_q^T)$. Some "improvements" of the ordinary LS method could be mentioned. For example, in the case of linear regression, when model errors are heteroscedastic, the so-called Feasible Generalized Least Squares (or Quasi-generalized Least Squares) estimation is asymptotically more accurate (see, e.g., Hamilton [43]). In the latter case, the main difference appears to be the definition of the estimator which is $\theta_n^* = (X^T\hat{\Omega}X)^{-1}X^T\hat{\Omega}Y$, with $\hat{\Omega} = \mathrm{diag}(\sigma_1^{-4}(\widehat{\theta}_1), ..., \sigma_n^{-4}(\widehat{\theta}_1))$. Another important aspect is that ordinary LS estimation can produce negative estimates of volatility – in order to avoid this problem the Constrained Least Squares estimation is used.

Next we present some basic aspects related to Quasi-maximum likelihood estimation (QMLE), which is without a doubt one of the most popular choices for parameter estimation of time series models such as GARCH, ARCH and others, including those with a long memory property. The name of this method entails "quasi", because the likelihood function we are maximizing to find the estimates of model parameters is written under the assumption of normally distributed innovations of the process. As it turns out, such an assumption is not critical for the asymptotic behavior of the estimator.

Let us now turn to the GARCH process to illustrate the main idea of QMLE. A number of papers consider the QMLE for GARCH processes, see, e.g., Hall and Yao [42], Francq and Zakoian [26], Berkes *et al.* [8], Berkes and Horváth [6], Berkes and Horváth [7]. The process we consider is a strictly stationary solution of equations

$$r_k = \sigma_k \zeta_k, \quad \sigma_k^2 = \omega_0 + \sum_{j=1}^{q} a_{0j} r_{k-j}^2 + \sum_{j=1}^{p} b_{0j} \sigma_{k-j}^2, \quad k \in \mathbb{Z}. \tag{2.7}$$

For estimation purposes, we assume that orders p and q are known. The true (unknown) parameters of this model are

$$\theta_0 = (\theta_{0,1}, ..., \theta_{0,p+q+1})^T = (\omega_0, a_{01}, ..., a_{0q}, b_{01}, ..., b_{0p})^T.$$

We want to estimate parameter θ_0 from the available realization of (2.7): $r_1, ..., r_n$, choosing the initial values of $r_0, ..., r_{1-q}, \tilde{\sigma}_0^2, ..., \tilde{\sigma}_{1-p}^2$. One of the possible choices of initial values is, for example, $r_0^2 = ... = r_{1-q}^2 = \tilde{\sigma}_0^2 = ... = \tilde{\sigma}_{1-p}^2 = r_1^2$. Since we start the process from chosen initial values, which affects the stationarity, further we work with $\left\{\tilde{\sigma}_t^2\right\}$. The QML estimator of $\widehat{\theta}_n$ is defined as

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L_n(\theta),$$

where $\Theta \subset (0, \infty) \times [0, \infty)^{p+q}$ is the parameter space, the quasi-likelihood function $L_n(\theta)$ is

$$L_n(\theta) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2}} \exp\left(-\frac{r_t^2}{2\tilde{\sigma}_t^2}\right).$$

The maximization problem can be equivalently rewritten to

$$\widehat{\theta}_n = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \left(\frac{r_t^2}{\tilde{\sigma}_t^2} + \log \tilde{\sigma}_t^2\right).$$

If the set of specific conditions for model coefficients a_j, b_j, and innovations ζ_k (e.g., $\mathrm{E}\zeta_k^4 < \infty$) is satisfied, the estimator is proved to be consistent and asymptotically normal. We intentionally do not go into detail in terms of these conditions and turn to the case of models that depend on infinite past.

Robinson and Zaffaroni [63] investigated the QMLE of ARCH(∞) models

$$r_k = \sigma_k \zeta_k, \quad \sigma_k^2 = \omega_0 + \sum_{j=1}^{\infty} \psi_{0j} r_{k-j}^2, \quad k \in \mathbb{Z}, \tag{2.8}$$

with

$$\omega_0 > 0, \quad \psi_{0j} > 0, \quad j \geq 1, \quad \sum_{j=1}^{\infty} \psi_{0j} < \infty.$$

It is the parametric version of ARCH(∞) as functions $\psi_j(\lambda)$ are assumed to be known and depend on vector $\lambda \in \mathbb{R}^r, r < \infty$, such that for the "true" value of $\lambda = \lambda_0$,

$$\psi_j(\lambda_0) = \psi_{0j}, \quad j \geq 1.$$

Note that they assume the strictly positive intercept of the model, i.e. $\omega_0 > 0$ (see Section 3 of this dissertation for more details on the ARCH(∞) process). In the context of infinite order ARCH-type models, it is common to define two likelihood functions: one which depends on infinite past

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{r_t^2}{\sigma_t^2(\theta)} + \log \sigma_t^2(\theta) \right), \quad 1 \leq t \leq n,$$

and another (more realistic) which depends on finite past

$$\widetilde{L}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{r_t^2}{\widetilde{\sigma}_t^2(\theta)} + \log \widetilde{\sigma}_t^2(\theta) \right), \quad 1 \leq t \leq n,$$

where

$$\widetilde{\sigma}_t^2 = \omega + \sum_{j=1}^{t-1} \psi_j(\lambda) r_{t-j}^2, \quad t \geq 1.$$

Accordingly, two estimators are considered:

$$\widehat{\theta}_n = \underset{\theta \in \Theta}{\arg\min} L_n(\theta), \quad \widetilde{\theta}_n = \underset{\theta \in \Theta}{\arg\min} \widetilde{L}_n(\theta). \tag{2.9}$$

Under the set of specific conditions, Robinson and Zaffaroni [63] prove the strong consistency and asymptotic normality of quasi-maximum likelihood estimators in (2.9).

It is worth mentioning a few words about the QML estimation for the class of linear ARCH (LARCH) models. In terms of the infinite order, these have a form

$$r_k = \sigma_k \zeta_k, \quad \sigma_k = \omega_0 + \sum_{j=1}^{\infty} b_j r_{k-j}, \tag{2.10}$$

where $\{\zeta_k, k \in \mathbb{Z}\}$ is a sequence of i.i.d. noise with zero mean and unit variance. The LARCH model can capture leverage effect and allow long memory modeling. However, volatility in (2.10) may assume negative and zero values, which not only limits the intuitive interpretation of σ_t as volatility, but also complicates the standard QML estimation of parameters in (2.10), because σ_k^{-2} and its derivatives may become arbitrarily small. As a result, the QML estimator for the LARCH model is, in general, inconsistent (for the finite order LARCH(q), see Francq and Zakoian [29]). As discussed in Section 4 of this dissertation, modified QMLE was proposed for the LARCH model by Beran and Schützner [5].

There are many other types of estimation methods which are beyond the scope of this dissertation. For example, some estimators are related to the spectral domain of the process – a perfect example is the Whittle estimation, often used in practice, which also covers long memory processes and was first introduced by Whittle [70]. Recall that in QMLE we deal with an objective function which includes the available observed values of the process and the volatility of some specific form. Whittle estimation optimizes the objective function, which is written in terms of spectral density and periodogram.

We are mainly interested in the QML estimation for the wide class of quadratic ARCH models with long memory; this is discussed in Section 4.

Chapter 3

Stationary integrated ARCH(∞) and AR(∞) processes with finite variance

In this chapter, we prove the long standing conjecture of Ding and Granger (1996, [20]) about the existence of the stationary Long Memory ARCH model with the finite fourth moment. This result follows from the necessary and sufficient conditions for the existence of covariance stationary integrated AR(∞), ARCH(∞) and FIGARCH models obtained in the present dissertation. We also prove that such processes always have long memory.

3.1 Introduction

As stated in Definition 2.3, a nonnegative random process $\{\tau_k\} = \{\tau_k,\ k \in \mathbb{Z}\}$ is said to satisfy an ARCH(∞) equation if there exists a sequence of nonnegative i.i.d. random variables $\{\varepsilon_k, k \in \mathbb{Z}\}$ with unit mean $\mathrm{E}\varepsilon_0 = 1$, a nonnegative number $\omega \geq 0$ and a deterministic sequence $b_j \geq 0$, $j = 1, 2, \ldots$, such that

$$\tau_k \;=\; \varepsilon_k \left(\omega + \sum_{j=1}^{\infty} b_j \tau_{k-j} \right), \qquad k \in \mathbb{Z}. \tag{3.1}$$

Unless stated otherwise, we assume that the process in (3.1) is causal, that is, for any k, τ_k can be represented as a measurable function $f(\varepsilon_k, \varepsilon_{k-1}, ...)$ of the present and past values $\varepsilon_s, s \leq k$ (see also Definition 2.4). Causality implies that a stationary process $\{\tau_k, k \in \mathbb{Z}\}$ is ergodic, and ε_k is independent of $\tau_s, s < k$. Therefore (and because $\mathrm{E}\varepsilon_0 = 1$),

$$E[\tau_k | \tau_s, s < k] = \sigma_k^2, \quad \sigma_k^2 = \omega + \sum_{j=1}^{\infty} b_j \tau_{k-j}.$$

A typical example of τ_k and ε_k in financial econometrics is squared returns and squared innovations, viz., $\tau_k = r_k^2$, $\varepsilon_k = \zeta_k^2$, where the return process $\{r_k, k \in \mathbb{Z}\}$ satisfies the ARCH(∞) equations

$$r_k = \zeta_k \sigma_k, \quad \sigma_k^2 = \omega + \sum_{j=1}^{\infty} b_j r_{k-j}^2 \qquad k \in \mathbb{Z}, \tag{3.2}$$

$\{\zeta_k, k \in \mathbb{Z}\}$ is a standardized i.i.d. $(0, 1)$-noise and σ_k is volatility. In this context, σ_k^2 is a conditional variance of returns r_k. The class of ARCH(∞) processes (3.1) includes the parametric stationary ARCH and GARCH models of Engle [24] and Bollerslev [10], where $r_k = \zeta_k \sigma_k$, and conditional variance σ_k^2 has the form

$$\sigma_k^2 = \omega + \sum_{j=1}^{q} \alpha_j r_{k-j}^2, \qquad k \in \mathbb{Z},$$

in case of ARCH(q) (taking $\alpha_j = b_j, j = 1, ..., q$, and $b_j = 0, j > q$), and

$$\sigma_k^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j r_{k-j}^2 + \sum_{j=1}^{p} \beta_j \sigma_{k-j}^2, \qquad k \in \mathbb{Z}, \tag{3.3}$$

in case of GARCH(p, q), where $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0, i = 1, 2, \ldots$. Equation (3.3) can be written as

$$\sigma_k^2 = \alpha_0 + \alpha(L) r_k^2 + \beta(L) \sigma_k^2,$$

where $\alpha(L) = \alpha_1 L + \cdots + \alpha_q L^q$ and $\beta(L) = \beta_1 L + \cdots + \beta_p L^p$. Now the expression

$$\sigma_k^2 = (1 - \beta(1))^{-1} \alpha_0 + (1 - \beta(L))^{-1} \alpha(L) r_k^2 \tag{3.4}$$

corresponds to ARCH(∞) equation (3.1) with $\omega = (1 - \beta(1))^{-1} \alpha_0$, and coefficients b_j are defined by $\sum_{j=1}^{\infty} b_j z^j = \alpha(z)/(1 - \beta(z))$. Kazakevičius and Leipus [51] proved that each strictly stationary solution of equations $r_k = \zeta_k \sigma_k$, with σ_k^2 as in (3.4), satisfies the associated ARCH(∞) equations.

The ARCH(∞) process was introduced by Robinson [62] in the context of hypothesis testing, and was considered as a class of parametric alternatives in testing serial correlation of disturbances in the static linear regression. Later, the ARCH(∞) process was studied by Kokoszka and Leipus [49] (change-point estimation in (3.2)), Giraitis *et al.* [31] (existence of stationary solution, its representation as a Volterra series, decay of covariance function, etc.), Giraitis and Surgailis [30] (bilinear equations: stationary solution, its covariance structure and long-memory properties; particular case of bilinear equations is the ARCH(∞) process), Leipus and Kazakevičius [51] (conditions for the existence of strictly stationary solution without moment conditions were obtained, as a generalization of results by Nelson [60] and Bougerol and Picard [12] for parametric ARCH and GARCH models), etc.

In contrast to the standard stationary GARCH(p, q) process whose autocorrelations decay exponentially:

$$\mathrm{corr}(r_0^2, r_k^2) = C \left(\sum_{j=0}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j \right)^k,$$

with coefficients $\alpha_j, \beta_j, \sum_{j=0}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1$, from (3.3) and a constant C independent of lag k, the ARCH(∞) process may have autocovariances $\mathrm{cov}(\tau_0, \tau_k)$ decaying to zero at a slower rate $k^{-\gamma}$, with $\gamma > 1$ arbitrarily close to 1. However, despite the possibility of a slow decay of autocovariances, a finite variance stationary solution to the ARCH equations in (3.1) with $\omega > 0$, if exists, has *short memory* or an *absolutely summable autocovariance function*, see Giraitis and Surgailis [30]. The existence of such a solution necessarily implies $\sum_{j=1}^{\infty} b_j < 1$ by $E\tau_k = \omega + (\sum_{j=1}^{\infty} b_j) E\tau_k > (\sum_{j=1}^{\infty} b_j) E\tau_k$, excluding stationary Integrated ARCH (IARCH) models with $\sum_{j=1}^{\infty} b_j = 1$. Because of the well-known phenomenon of long memory of squared returns, the latter finding may be considered a limitation to ARCH modeling. Subsequently, it initiated and justified the study of other ARCH-type models, for which the long memory property can be rigorously established (see, e.g., Giraitis, Robinson and Surgailis [32], where they considered the

Linear ARCH (LARCH) model with $\sigma_k = \omega + \sum_{j=1}^{\infty} b_j r_{k-j}$, and Giraitis, Leipus and Surgailis [35]).

A particular case of the IARCH model is the well-known FIGARCH (Fractionally Integrated GARCH) equation

$$\tau_k \;=\; \varepsilon_k \big\{\omega + \big(1-(1-L)^d\big)\tau_k\big\} \;=\; \varepsilon_k \left(\omega + \sum_{j=1}^{\infty} b_j \tau_{k-j}\right), \quad k \in \mathbb{Z}, \quad (3.5)$$

where $0 < d < 1/2$ is the fractional differencing parameter, L is the backshift operator and coefficients b_j are determined by the generating function $B(z) = \sum_{j=1}^{\infty} b_j z^j = 1-(1-z)^d$. Here, $b_j > 0$, $\sum_{j=1}^{\infty} b_j = 1$, and $b_j = O(j^{-1-d})$ decay hyperbolically with $j \to \infty$. The FIGARCH equation was introduced by Baillie, Bollerslev, and Mikkelsen [3] to capture the long memory effect in volatility. Independently of the last paper, Ding and Granger [20] introduced the LM(d)-ARCH model

$$r_k^2 = \zeta_k^2 \sigma_k^2, \quad \sigma_k^2 = \mu(1-\theta) + \theta\left(1-(1-L)^d\right) r_k^2, \qquad k \in \mathbb{Z}, \qquad (3.6)$$

where $\theta \in [0,1], \mu > 0$, and r_k, ζ_k are related to τ_k, ε_k as in (3.2). A similar long memory model for absolute returns was proposed by Granger and Ding [39]. Ding and Granger [20] derived (3.6) via contemporaneous aggregation of a large number of GARCH(1,1) processes with random Beta distributed coefficients. Ding and Granger [20] note that in the integrated case $\theta = 1$, (3.6) coincides with the special case $\omega = 0$ of the FIGARCH model in (3.5). Ding and Granger [20], p. 206–207, argue that a stationary solution of (3.6) with the finite fourth moment has long memory, in the sense that

$$\mathrm{corr}(r_0^2, r_k^2) \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{-1+2d}. \qquad (3.7)$$

The results in Baillie *et al.* [3] imply a similar long memory behavior of the FIGARCH model. However, the existence of the stationary solution of the LM(d)-ARCH equation in (3.6) with the finite fourth moment was not rigorously established and the validity of (3.7) remained open. See Davidson [18], Giraitis *et al.* [31], Kazakevičius and Leipus [16], Mikosch and Stărică ([57], [58]) for a discussion of controversies surrounding FIGARCH and LM(d)-ARCH models. For example, Davidson [18], using the findings by Giraitis *et al.* [31], Kazakevičius and Leipus [15], suggests that, in general, the FIGARCH process should not be treated as a "long memory" process but instead as a "hyperbolic memory" process. Mikosch and Stărică [57] emphasized that although the FIGARCH model is often mentioned in literature on long memory econometrics, an important drawback is that rigorous proof of the existence of a stationary version of the FIGARCH process is not available.

In the present dissertation we solve the long standing conjecture (3.7) of Ding and Granger [20]. We prove that the necessary and sufficient condition for the existence of a covariance stationary solution of the FIGARCH equation in (3.5) with $\omega = 0$ is

$$E\varepsilon_0^2 < \frac{\Gamma(1-2d)}{\Gamma(1-2d) - \Gamma^2(1-d)}, \tag{3.8}$$

and, therefore, conditions (3.8) and $\theta = 1$ are necessary and sufficient for (3.7)[1]. See Corollary 3.2 below.

The above-mentioned result is a particular case of a more general

[1] Condition (3.8) for the existence of a stationary solution of the FIGARCH equation in (3.5) with $\omega = 0$ was independently obtained in the unpublished paper by Koulikov [50] who used a similar approach for constructing the solution. However, proof in Koulikov ([50], Theorem 2) is based on erroneous assumption (9), which contradicts the IARCH condition $\sum_{j=1}^{\infty} b_j = 1$.

result concerning the integrated ARCH(∞), or IARCH(∞), equation with zero intercept:

$$\tau_k = \varepsilon_k \left(\sum_{j=1}^{\infty} b_j \tau_{k-j} \right), \quad k \in \mathbb{Z}, \quad \text{with} \quad \sum_{j=1}^{\infty} b_j = 1. \tag{3.9}$$

Note that for $\sum_{j=1}^{\infty} b_j < 1$, equation (3.9) only has a trivial stationary solution $\tau_k \equiv 0$ with finite mean, which follows from $E\tau_k = (\sum_{j=1}^{\infty} b_j) E\tau_k$, by taking expectations. Our main result is Theorem 3.1, stating that, in addition to the zero solution, a nontrivial covariance stationary solution of the IARCH equation in (3.9) with $b_j \geq 0$ exists if and only if

$$\|g\|^2 = \sum_{j=0}^{\infty} g_j^2 < (1+\sigma^2)/\sigma^2, \tag{3.10}$$

where $\sigma^2 = \operatorname{Var}(\varepsilon_0)$ and coefficients g_j are determined from the power expansion

$$\sum_{j=0}^{\infty} g_j L^j = (1 - B(L))^{-1}, \qquad \text{where} \qquad B(L) = \sum_{j=1}^{\infty} b_j L^j. \tag{3.11}$$

Condition (3.10) rules out integrated GARCH(p, q) as well as any integrated ARCH(∞) models with sufficiently fast decaying lags which are known to admit a stationary solution with infinite variance, see Kazakevičius and Leipus [16], Douc *et al.* [21], Robinson and Zaffaroni [63]. It turns out that covariance stationary solutions of (3.9) always have long memory, in the sense that the covariance function is nonsummable and the spectral density is infinite at the origin, see Corollary 3.1.

The main idea of constructing a stationary L^2-solution (i.e. whose second moment is finite and series in (3.9) converges in mean square) τ_k

of the IARCH equation (3.9) with mean $\mu = E\tau_k > 0$ is the reduction of equation (3.9) to the linear Integrated AR (IAR) equation for the centered process $Y_k = \tau_k - \mu$:

$$Y_k \;=\; \sum_{j=1}^{\infty} b_j Y_{k-j} + z_k, \qquad k \in \mathbb{Z}, \tag{3.12}$$

with a conditionally heteroskedastic martingale difference noise $\{z_k, k \in \mathbb{Z}\}$ defined as

$$z_k \;=\; \zeta_k \left(\mu\sigma + \sigma \sum_{j=1}^{\infty} b_j Y_{k-j} \right), \tag{3.13}$$

where $\zeta_k = (\varepsilon_k - 1)/\sigma$, $\sigma^2 = \mathrm{Var}(\varepsilon_0) < \infty$. In turn, based on (3.12) and (3.13), the process $\{z_k, k \in \mathbb{Z}\}$ can be defined as a stationary solution of the LARCH (Linear ARCH) equation (3.18) with standardized zero mean i.i.d. innovations $\{\zeta_k, k \in \mathbb{Z}\}$ discussed in Giraitis *et al.* ([32], [33]), given by convergent Volterra series in (3.19). Then, a causal L^2-solution $\{Y_k, k \in \mathbb{Z}\}$ can be obtained by inverting the linear IAR equation in (3.12).

The last question is tackled in Section 3.3, where we establish sufficient and necessary conditions for the existence of a covariance stationary solution of the linear Integrated AR(∞) equation generalizing (3.12):

$$x_k - \sum_{j=1}^{\infty} b_j x_{k-j} = \xi_k, \qquad k \in \mathbb{Z}, \tag{3.14}$$

where $b_j \geq 0$, $\sum_{j=1}^{\infty} b_j = 1$, and $\{\xi_k, k \in \mathbb{Z}\}$ is a stationary short memory process, in particular, white noise. Theorem 3.2 states that covariance stationary solutions of (3.14) always have long memory, which originates from integration property $\sum_{j=1}^{\infty} b_j = 1$ with an infinite number of $b_j \geq 0$.

This result is in deep contrast with the well-known fact that integrated AR$(p), p < \infty$, processes are nonstationary and need to be differenced to achieve stationarity.

Section 3.2 discusses stationary L^2-solutions of ARCH(∞) (3.1) and bilinear equations (3.12)–(3.13) and their mutual relationship. It contains Theorem 3.1 together with several corollaries. Section 3.3 discusses solvability and second-order properties of IAR(∞) equation (3.14). All proofs are relegated to Sections 3.4 and 3.5.

3.2 Stationary solutions of FIGARCH, IARCH and ARCH equations

In this section, we discuss the existence of a stationary L^2-solution of ARCH(∞) equation (3.1) in the integrated case $\sum_{j=1}^{\infty} b_j = 1$. We first explain the idea of solving ARCH(∞) equation (3.1) with a nonnegative i.i.d. noise $\{\varepsilon_k, k \in \mathbb{Z}\}$ by reducing it to a bilinear equation with a zero mean i.i.d. noise $\{\zeta_k, k \in \mathbb{Z}\}$ used by Giraitis and Surgailis [30]. Recall the definition of the ARCH(∞) model in (3.1). Specifically, for a stationary ARCH(∞) process τ_k in (3.1) with mean $E\tau_k = \mu$, we set

$$Y_k = \tau_k - \mu.$$

Let $\theta = \sum_{j=1}^{\infty} b_j$. We focus on two cases: a) $\omega > 0$ and $0 < \theta < 1$, and b) $\omega = 0$ and $\theta = 1$. As noted above, the case $\omega = 0$ and $\theta < 1$ is not of particular interest and is excluded from the subsequent discussion since it leads to a unique trivial solution $\tau_k \equiv 0$. By taking expectations, equation (3.1) implies $E\tau_k = \omega + \theta E\tau_k$, or $\mu = E\tau_k = \omega/(1-\theta)$ in case a), while in case b), it does not contradict a free choice of $\mu > 0$. Motivated by these facts, put

$$\mu = \begin{cases} \omega/(1-\theta), & \text{if } \theta < 1 \text{ and } \omega > 0, \\ \text{any positive number } \mu > 0, & \text{if } \theta = 1 \text{ and } \omega = 0. \end{cases}$$

Assume $\sigma^2 = \mathrm{Var}(\varepsilon_0) < \infty$ and let $\{\zeta_k = (\varepsilon_k - 1)/\sigma,\ k \in \mathbb{Z}\}$ be the centered i.i.d. noise (recall that ε_k in (3.1) are standardized: $E\varepsilon_k = 1$). With this notation, the ARCH equation of (3.1) can be written as the bilinear equation

$$Y_k = \sum_{j=1}^{\infty} b_j Y_{k-j} + \zeta_k \left(\mu\sigma + \sigma \sum_{j=1}^{\infty} b_j Y_{k-j} \right), \tag{3.15}$$

see also Giraitis and Surgailis [30]. As noted by Giraitis *et al.* [32], Giraitis and Surgailis [30], (3.15) is different from bilinear equations discussed by Granger and Andersen [38], Subba Rao [61] due to the presence of cross terms $\zeta_k Y_{k-j}$. Let

$$z_k = Y_k - \sum_{j=1}^{\infty} b_j Y_{k-j} = (1 - B(L))Y_k.$$

Then $Y_k = (1 - B(L))^{-1} z_k = G(L) z_k = \sum_{j=0}^{\infty} g_j z_{k-j}$, and

$$\sigma \sum_{j=1}^{\infty} b_j Y_{k-j} = \sigma B(L)(1 - B(L))^{-1} z_k = H(L) z_k = \sum_{j=1}^{\infty} h_j z_{k-j},$$

where coefficients g_j, h_j, of the generating functions $G(z), H(z)$ are defined by

$$G(z) = \frac{1}{1-B(z)} = \sum_{j=0}^{\infty} g_j z^j, \quad H(z) = \frac{\sigma B(z)}{1-B(z)} = \sum_{j=1}^{\infty} h_j z^j, \quad |z| < 1. \tag{3.16}$$

Notice that $h_j = \sigma g_j \, (j \geq 1)$, $g_0 = 1$, $h_0 = 0$, follows from equality $H(z) = \sigma(G(z) - 1)$, which, in turn, follows from (3.16). Hence (3.15) can be written as the system of two equations:

$$(a)\ Y_k = \sum_{j=1}^{\infty} b_j Y_{k-j} + z_k, \quad (b)\ z_k = \zeta_k \left(\mu\sigma + \sum_{j=1}^{\infty} h_j z_{k-j} \right). \tag{3.17}$$

Note that equation (3.17)(b) does not contain Y_k and coincides with the so-called LARCH model studied by Giraitis *et al.* ([32], [33]) and elsewhere. Also observe that $\{z_k, k \in \mathbb{Z}\}$ is a martingale difference sequence which can be written as

$$z_k = \zeta_k v_k, \qquad v_k = \mu\sigma + \sum_{j=1}^{\infty} h_j z_{k-j}, \tag{3.18}$$

where v_k may be interpreted as volatility. A stationary solution $\{z_k, k \in \mathbb{Z}\}$ of equation (3.18) is constructed in terms of causal Volterra series in i.i.d. innovations $\zeta_s, s \leq k$:

$$z_k = \mu\sigma\zeta_k \left(1 + \sum_{m=1}^{\infty} \sum_{s_m < \cdots < s_1 < k} h_{k-s_1} h_{s_1-s_2} \cdots h_{s_{m-1}-s_m} \zeta_{s_1} \cdots \zeta_{s_m} \right), \tag{3.19}$$

see Giraitis *et al.* ([32], [33]). The series in (3.19) converges in L^2 if and only if

$$\begin{aligned} \sigma_z^2 = Ez_k^2 &= (\mu\sigma)^2 \left(1 + \sum_{m=1}^{\infty} \sum_{s_m < \cdots < s_1 < k} h_{k-s_1}^2 h_{s_1-s_2}^2 \cdots h_{s_{m-1}-s_m}^2\right) \\ &= (\mu\sigma^2) \left(1 + \sum_{m=1}^{\infty} \|h\|^{2m}\right) < \infty, \end{aligned} \tag{3.20}$$

or $\|h\|^2 = \sum_{j=1}^{\infty} h_j^2 < 1$, which is equivalent to

$$\|g\|^2 = \sum_{j=0}^{\infty} g_j^2 < (1+\sigma^2)/\sigma^2.$$

After solving equation (3.17)(b), equation (3.17)(a) in the integrated case $\theta = \sum_{j=1}^{\infty} b_j = 1$ represents a particular case of the IAR(∞) model with causal uncorrelated noise $\{z_k\}$ discussed in Theorem 3.2 below. Accordingly, the stationary solution of bilinear equation (3.15) and, consequently, of ARCH equation (3.1) can be obtained by inverting (3.17)(a), that is,

$$\begin{aligned} Y_k &= (1 - B(L))^{-1} z_k = \sum_{j=0}^{\infty} g_j z_{k-j} \qquad (3.21) \\ &= \mu\sigma \left(\sum_{m=1}^{\infty} \sum_{-\infty < s_m < \cdots < s_1 \le k} g_{k-s_1} h_{s_1-s_2} \cdots h_{s_{m-1}-s_m} \zeta_{s_1} \cdots \zeta_{s_m} \right), \end{aligned}$$

as a solution of the AR(∞) equation with martingale difference innovations z_{k-j} determined by equation (3.17)(b), or (3.18), see Proposition 3.1 (iii).

In what follows, the term "causal" indicates a stationary process $\{y_k, k \in \mathbb{Z}\}$ written as a measurable function of present and past values $\zeta_s, s \le k$, or, equivalently, $\varepsilon_s, s \le k$ (see also Definition 2.4).

Definition 3.1. *By an L^2-solution of equations (3.1), (3.15), (3.17), we mean a random process with the finite second moment such that all series in these equations converge in mean square and the corresponding equations hold for each $k \in \mathbb{Z}$.*

The main result of this chapter and one of the most important findings in this dissertation overall, is the following theorem, which establishes sufficient and necessary conditions for the existence of a causal L^2-solution $\{\tau_k, k \in \mathbb{Z}\}$ of ARCH(∞) equation (3.1) and $\{(Y_k, z_k, k \in \mathbb{Z})\}$ of bilinear equations in (3.15), (3.17). Denote the transfer function (see, e.g., Definition 2.2)

$$A(x) = (1 - B(\mathrm{e}^{\mathrm{i}x}))^{-1}, \quad B(\mathrm{e}^{\mathrm{i}x}) = \sum_{j=1}^{\infty} b_j \mathrm{e}^{\mathrm{i}jx}, \qquad x \in \Pi := [-\pi, \pi],$$

and set $\|g\|^2 = \sum_{j=0}^{\infty} g_j^2$ and $\|A\|^2 = \int_\Pi |A(x)|^2 \mathrm{d}x$.

Theorem 3.1. *Let $\omega \geq 0$, $0 < \theta \leq 1$, excluding the case $\omega = 0, 0 < \theta < 1$.*

(a) *ARCH equation (3.1) has a nontrivial causal L^2-solution $\{\tau_k, k \in \mathbb{Z}\}$ if and only if*

$$\|g\|^2 < (1 + \sigma^2)/\sigma^2. \tag{3.22}$$

Condition (3.22) is equivalent to

$$\|A\|^2 < 2\pi(1 + \sigma^2)/\sigma^2. \tag{3.23}$$

(b) *Let (3.22) or (3.23) be satisfied, and let Y_k be defined as in (3.21), (3.19).*

(i) *If $\omega > 0$, $0 < \theta < 1$, then ARCH equation (3.1) has a unique causal L^2-solution $\{\tau_k = \mu + Y_k, k \in \mathbb{Z}\}$, where $\mu = \omega/(1 - \theta) = E\tau_k$.*

(ii) *If* $\omega = 0$, $\theta = 1$, *then for each* $\mu > 0$, $\{\tau_k = \mu + Y_k, k \in \mathbb{Z}\}$ *is a unique causal* L^2*-solution of (3.1) with mean* $E\tau_k = \mu$.

Theorem 3.1 is new only in the integrated case $\theta = 1$, since for $\theta < 1$ it follows from the paper of Giraitis and Surgailis [30]. The case $\theta < 1$ is included above for comparison. While for $\theta < 1$ the solution is unique, for $\theta = 1$ IARCH equation (3.9) has an infinite number of causal L^2-solutions parametrized by $E\tau_k = \mu$. Since coefficients g_j are expressed through b_j via multiple infinite series, see (3.27), direct verification of condition (3.22) may be difficult. On the other hand, condition (3.23) in some cases can be verified rather easily if the transfer function $A(x)$ is explicitly known, as in the case of the FIGARCH model.

The following corollary establishes the long memory property of the stationary IARCH model.

Corollary 3.1. *IARCH equation (3.9) has a nontrivial stationary causal* L^2*-solution if and only if* $\sigma^2 = \mathrm{Var}(\varepsilon_0)$ *and* b_j *satisfy condition (3.23) (or, equivalently, (3.22)). In the latter case,*

(i) *for each* $\mu > 0$, *the process* $\{\tau_k = \mu + Y_k, k \in \mathbb{Z}\}$ *with* Y_k *defined in (3.21), (3.19), is a unique causal* L^2*-solution of (3.9) with mean* $E\tau_k = \mu$.

(ii) *the covariance function of the solution* $\{\tau_k = \mu + Y_k, k \in \mathbb{Z}\}$ *is given by*

$$\mathrm{cov}(\tau_0, \tau_k) = \sigma_z^2 \sum_{j=0}^{\infty} g_j g_{k+j}, \tag{3.24}$$

where σ_z^2 *is given in (3.20).*

(iii) *the covariance function in (3.24) is nonnegative,* $\mathrm{cov}(\tau_0, \tau_k) \geq 0$, *and nonsummable:* $\sum_{k\in\mathbb{Z}} \mathrm{cov}(\tau_0, \tau_k) = \infty$. *Moreover,* $\{\tau_k, k \in \mathbb{Z}\}$ *has spectral* *density*

$$f(x) = \mu^2(\sigma_z^2/2\pi)|1 - B(\mathrm{e}^{-\mathrm{i}x})|^{-2}, \quad x \in \Pi,$$

that is unbounded at zero frequency: $f(x) \to \infty$, *as* $x \to 0$.

Corollary 3.1 together with Lemma 3.1 (iii) imply that the IARCH model in (3.9) with $\omega = 0$ does not have a stationary solution with finite variance if b_j tend to zero fast enough, for example, exponentially, or decay at rate $b_j = O(j^{-\gamma})$, for some $\gamma \geq 3/2$. In contrast, sufficient conditions for the existence of a stationary IARCH process with nonzero intercept $\omega > 0$ and infinite mean $E\tau_k = \infty$, obtained in Kazakevičius and Leipus [16], and Douc *et al.* [21], require an exponential decay of b_j, as $j \to \infty$.

The following corollary details the case of the FIGARCH equation in (3.5) with zero intercept $\omega = 0$. It establishes the existence of stationary long memory FIGARCH processes $\{\tau_k, k \in \mathbb{Z}\}$ and shows that their covariance function $\mathrm{cov}(\tau_k, \tau_0)$ decays to zero hyperbolically slowly as in (3.25).

Corollary 3.2. *For the FIGARCH model in (3.5) with* $\omega = 0$ *and* $d \in (0, 1/2)$, *condition (3.22) is equivalent to (3.8), that is,*

$$\mathrm{E}\varepsilon_0^2 < \frac{\Gamma(1-2d)}{\Gamma(1-2d) - \Gamma^2(1-d)}.$$

Under this condition, the statements of Corollary 3.1 hold. Moreover, as $k \to \infty$, *the covariance and spectral density of the FIGARCH process* $\{\tau_k, k \in \mathbb{Z}\}$ *with* $E\tau_k = \mu$ *satisfy*

$$\mathrm{cov}(\tau_0, \tau_k) \sim \mu^2 c_\gamma k^{-1+2d}, \tag{3.25}$$

$$f(x) = (\sigma_z^2/2\pi)|1 - \mathrm{e}^{\mathrm{i}x}|^{-2d} \sim (\sigma_z^2/2\pi)|x|^{-2d}, \quad x \to 0.$$

where $c_\gamma = \sigma_z^2 \Gamma(1-2d)/\{\Gamma(d)\Gamma(1-d)\}$, *and*

$$\sigma_z^2 = \sigma^2 / \big(1 + \sigma^2 - \sigma^2 (\Gamma(1-2d)/\Gamma^2(1-d))\big).$$

For comparison, Corollary 3.3 below recovers the results on the existence of a stationary finite variance solution of the ARCH(∞) equation with $\theta = \sum_{j=1}^{\infty} b_j < 1$, obtained by Giraitis and Surgailis [30]. As noted above, the existence of such a solution in this case necessarily implies $E\tau_k = \mu = \omega/(1-\theta)$. In sharp contrast to a finite variance stationary IARCH process, which can only have long memory, see Corollary 3.1, the stationary finite variance ARCH process with $\theta < 1$ always has short memory.

Corollary 3.3. *ARCH(∞) equation (3.1) with* $\omega > 0$ *and* $\theta = \sum_{j=1}^{\infty} b_j < 1$ *has a unique stationary causal* L^2*-solution* $\{\tau_k, k \in \mathbb{Z}\}$ *if and only if condition (3.22) is satisfied. The above solution is given by* $\{\tau_k = \mu + Y_k, k \in \mathbb{Z}\}$*, with* $\mu = \omega/(1-\theta)$*, and* Y_k *defined in (3.21), (3.19). It has mean* $E\tau_k = \mu = \omega/(1-\theta)$ *and a nonnegative covariance function given in (3.24). Moreover,*

$$\sum_{k=0}^{\infty} \mathrm{cov}(\tau_0, \tau_k) < \infty, \quad \sum_{k=0}^{\infty} g_k < \infty.$$

Corollary 3.4 discusses weak convergence in the Skorohod space $D[0,1]$, denoted by $\to_{D[0,1]}$, of the partial sums process of $\{\tau_k, k \in \mathbb{Z}\}$. Part (i) of this corollary is known, see Giraitis *et al.* ([34], [31]). Below, $\{B(t), t \in [0,1]\}$ denotes the standard Brownian motion with variance $EB^2(t) = t$ and $\{B_{d+1/2}(t), t \in [0,1]\}$ a fractional Brownian motion with variance $EB^2_{d+1/2}(t) = t^{2d+1}$, $d \in (0, 1/2)$ (see also Definitions 2.10 and 2.11).

Corollary 3.4. *Suppose that (3.22) holds.*

(i) *Let* $\omega > 0,\ \theta < 1$ *and* $\{\tau_k, k \in \mathbb{Z}\}$ *be the ARCH(∞) process as in Corollary 3.3. Then*

$$n^{-1/2}\sum_{k=1}^{[nt]}(\tau_k - E\tau_k) \to_{D[0,1]} s^2 B(t), \qquad s^2 = \sum_{k\in\mathbb{Z}} \mathrm{cov}(\tau_0, \tau_k).$$

(ii) *Let* $\{\tau_k, k \in \mathbb{Z}\}$ *be the FIGARCH process as in Corollary 3.2. Then*

$$n^{-1/2-d}\sum_{k=1}^{[nt]}(\tau_k - E\tau_k) \to_{D[0,1]} s_d B_{d+1/2}(t), \qquad s_d^2 = \mu^2 c_\gamma/(d(1+2d)).$$

We are able to give a final answer to conjecture (3.7) of Ding and Granger [20], which assumes the existence of a stationary solution $\{r_k, k \in \mathbb{Z}\}$ of the LM(d)-ARCH model in (3.6) with $Er_k^4 < \infty$, for arbitrary parameters $\theta \in (0,1], 0 < d < 1/2$, and $\mu > 0$. Although this conjecture is proved only for $\theta = 1$, the fact that it is invalid for all $0 < \theta < 1$ is also new, since previously the failure of (3.7) was only shown for $\theta < 1/\sqrt{E\zeta_0^4} < 1$, see Giraitis *et al.* [31], Section 4.

Corollary 3.5. *Conjecture (3.7) of Ding and Granger (1996) about the LM(d)-ARCH model in (3.6) is true if and only if* $\theta = 1$ *and* $E\zeta_0^4 = E\varepsilon_0^2$ *satisfy condition (3.8).*

3.3 Stationary Integrated AR(∞) processes: Origins of long memory

As explained in the previous two sections of this chapter, our construction of a stationary solution of the IARCH model relies on solving IAR equation

(3.12) with martingale difference innovations $\{z_k, k \in \mathbb{Z}\}$. In particular, we want to know which conditions on filter b_j guarantee that the IAR equation has a stationary solution and when does this solution have covariance long memory, in the sense that its covariance function is nonsummable.

It turns out that the two questions are closely related, in the sense that the existence of a stationary solution of the IAR equation implies the long memory property of its solution. This question is of independent interest apart from ARCH models, since it indicates a general mechanism for generating a long memory process, different from fractional differencing or the ARFIMA(p, d, q) model commonly used in time series literature (see, e.g., Brockwell and Davis [13], Giraitis, Koul, and Surgailis [36]). Being a technical tool for generating parametric long memory time series, fractional filtering/differencing cannot fully explain the phenomenon and how long memory is induced, which sometimes leads to controversies justifying the use of long memory processes and explaining the mechanism for generating them. See Lieberman and Phillips [53] for an illustrative analysis of how long memory may arise in realized volatility.

In this section, we discuss the stationary solution of the Integrated AR(∞) equation:

$$x_k - \sum_{j=1}^{\infty} b_j x_{k-j} = \xi_k, \qquad k \in \mathbb{Z}, \tag{3.26}$$

where b_j are nonnegative, $\sum_{j=1}^{\infty} b_j = 1$, and $\{\xi_k, k \in \mathbb{Z}\}$ is a white noise (a stationary sequence of uncorrelated random variables with zero mean and finite variance $\sigma_\xi^2 = E\xi_0^2 < \infty$). In this section, by stationarity we mean weak sense or covariance stationarity, since no other properties of random variables with exception of the two finite first moments will be used.

Definition 3.2. *We say that a random process $\{x_k, k \in \mathbb{Z}\}$ is a L^2-solution of (3.26) if $Ex_k^2 < \infty$ for each $k \in \mathbb{Z}$, the series $\sum_{j=1}^{\infty} b_j x_{k-j}$ converges in mean square, and (3.26) holds.*

The above definition is very general and does not assume causality or even ergodicity of $\{x_k, k \in \mathbb{Z}\}$ since any constant "random variable" $x \equiv x_k$, $Ex^2 < \infty$, is a L^2-solution of the homogeneous equation $x_k - \sum_{j=1}^{\infty} b_j x_{k-j} = 0$. As for IARCH equation (3.9), a (stationary) L^2-solution $\{x_k, k \in \mathbb{Z}\}$ of (3.26), if exists, is not unique: for any real μ, $\{x_k + \mu, k \in \mathbb{Z}\}$ is also a L^2-solution of (3.26). The existence of such a solution implies that b_j cannot vanish for j large enough, for example, a unit root model $x_k - x_{k-1} = \xi_k$ does not have a stationary solution.

A causal solution of (3.26) can be constructed by inverting the filter $1 - B(z)$ with inverse filter coefficients g_j, $j \geq 0$, as defined in (3.16), by using the power expansion of the analytic function $G(z) = (1 - B(z))^{-1} = \sum_{j=0}^{\infty} g_j z^j$ on the unit disc $\{|z| < 1\}$. The resulting coefficients are nonnegative and given by

$$g_j = \sum_{m=1}^{j} \sum_{0<s_{m-1}<\cdots<s_1<j} b_{j-s_1} b_{s_1-s_2} \cdots b_{s_{m-2}-s_{m-1}} b_{s_{m-1}}, \quad j \geq 1, \qquad g_0 = 1, \tag{3.27}$$

which follows from equality $(1 - B(z))^{-1} = \sum_{m=0}^{\infty} B^m(z)$. Assuming that $\|g\| = (\sum_{j=0}^{\infty} g_j^2)^{1/2} < \infty$, we can define a stationary L^2-solution of (3.26) as

$$\tilde{x}_k = \sum_{j=0}^{\infty} g_j \xi_{k-j}, \qquad k \in \mathbb{Z}. \tag{3.28}$$

As shown in Lemma 3.2 below, if the transfer function $A(x) = (1 -$

$B(\mathrm{e}^{\mathrm{i}x}))^{-1}$ is L^2-integrable: $\|A\| = (\int_\Pi |A(x)|^2 \mathrm{d}x)^{1/2} < \infty$, the Fourier coefficients of $A(x)$ agree with g_j in (3.27):

$$g_j \;=\; (2\pi)^{-1} \int_\Pi A(x)\mathrm{e}^{-\mathrm{i}xj}\mathrm{d}x \quad \text{and} \quad A(x) = \sum_{j=0}^{\infty} g_j \mathrm{e}^{\mathrm{i}xj}. \tag{3.29}$$

Notice that equalities (3.29) are not obvious since the g_js are defined by the power expansion of $G(z)$ in the open disc $|z| < 1$, while the definition of $A(x)$ requires only $B(\mathrm{e}^{\mathrm{i}x}) \neq 1$ a.e.

The next theorem establishes the equivalence of conditions $\|g\| < \infty$ and $\|A\| < \infty$ and representations (3.27) and (3.29). It also obtains conditions for the existence and uniqueness of a stationary L^2-solution of (3.26) and its long memory property.

Theorem 3.2. (i) *Assumption $\|g\| < \infty$ is necessary and sufficient for the existence of a stationary L^2-solution $\{x_k, k \in \mathbb{Z}\}$ of (3.26).*

(ii) *If $\|g\| < \infty$, then with $\tilde{x}_k$ as in (3.28) for each real μ,*

$$x_k = \mu + \tilde{x}_k, \quad k \in \mathbb{Z}, \tag{3.30}$$

is a stationary L^2-solution of (3.26) with $Ex_k = \mu$. The above solution is unique in the class of all stationary linear processes $x_k = \mu + \sum_{j\in\mathbb{Z}} c_j \xi_{k-j}$ with $\sum_{j\in\mathbb{Z}} c_j^2 < \infty$.

(iii) *The solution x_k in (3.30) has a nonnegative and nonsummable covariance function:*

$$\mathrm{cov}(x_0, x_k) = \sigma_\xi^2 \sum_{j=0}^{\infty} g_j g_{k+j} \geq 0, \qquad \sum_{k\in\mathbb{Z}} \mathrm{cov}(x_0, x_k) = \infty, \tag{3.31}$$

and unbounded spectral density $f(x) = \frac{\sigma_\xi^2}{2\pi}|1-B(\mathrm{e}^{\mathrm{i}x})|^{-2}$ *with* $\lim_{x\to 0} f(x) = \infty$.

(iv) $\|g\| < \infty$ *implies* $\|A\| < \infty$ *and (3.29). Conversely,* $\|A\| < \infty$ *implies* $\|g\| < \infty$.

A surprising consequence of Theorem 3.2 is the fact that stationary solution (3.28) of (3.26) does not exist if the b_js vanish for j large enough. The validity of this conclusion is not obvious from the representation of g_j in (3.27) but follows easily from (3.29). Indeed, since $|A(x)|^{-1} = |1 - B(\mathrm{e}^{\mathrm{i}x})| = |\sum_{j=0}^{\infty} b_j(1 - \mathrm{e}^{\mathrm{i}jx})| \le |x| \sum_{j=1}^{\infty} j|b_j| \le C|x|$, this implies $\int_\Pi |A(x)|^2 \mathrm{d}x \ge C^{-2} \int_\Pi x^{-2} \mathrm{d}x = \infty$ and $\|g\| = \infty$ according to (3.29). The above argument combined with Lemma 3.1 (iii) is formalized in the following corollary.

Corollary 3.6. *The IAR(∞) equation in (3.26) does not have a stationary L^2-solution if the b_js decay as $j^{-3/2}$ or faster. In particular, the latter holds if $b_j = 0, j > j_0$ for some $j_0 \ge 1$, or $b_j = O(\mathrm{e}^{-cj})$ for $j \ge 1, c > 0$.*

The requirement of Theorem 3.2 that the r.h.s. $\{\xi_k, k \in \mathbb{Z}\}$ in IAR equation (3.26) is white noise, is restrictive and can be relaxed. Theorem 3.3 extends Theorem 3.2 to the case when $\{\xi_k, k \in \mathbb{Z}\}$ is a short memory process as precised below.

Theorem 3.3. *Let $\{\xi_k, k \in \mathbb{Z}\}$ be a stationary process with zero mean, finite variance and a spectral density f_ξ which is bounded away from 0 and ∞:*

$$c_1 \le f_\xi(x) \le c_2, \quad \forall\, x \in \Pi, \qquad \exists\, 0 < c_1 < c_2 < \infty.$$

Then statements (i) and (ii) of Theorem 3.2 about a stationary solution of (3.26) remain valid, while statement (iii) has to be modified as follows:

(iii′) *The solution* $\{x_k, k \in \mathbb{Z}\}$ *in (3.30) has unbounded spectral density*

$$f(x) = |1 - B(\mathrm{e}^{\mathrm{i}x})|^{-2} f_\xi(x),$$

that satisfies $\lim_{x\to 0} f(x) = \infty$*, and a nonsummable autocovariance function:*

$$\sum_{k\in\mathbb{Z}} |\mathrm{cov}(x_0, x_k)| = \infty.$$

Apparently, the class of stationary IAR(∞) processes with long memory satisfying the conditions of Theorems 3.2 or 3.3 is quite large. Since condition $\theta = \sum_{j=1}^{\infty} b_j = 1$ does not assume any particular form of b_j, it seems that the spectral density of an IAR(∞) process need not grow regularly as a power function $|x|^{-\alpha}$, $0 < \alpha < 1$, at $x = 0$ and, similarly, the covariance function need not decay regularly with the lag as $k^{-1+\alpha}$. The latter properties are key features of fractionally integrated ARFIMA models (see, e.g., Hosking [44], also Giraitis *et al.* [36], Chapter 7).

Example 3.1. The ARFIMA($0, d, 0$) model is defined as a stationary solution of the equation

$$(1 - L)^d x_k = \xi_k, \qquad 0 < d < 1/2,$$

where $\{\xi_k, k \in \mathbb{Z}\}$ is uncorrelated white noise with $E\xi_k = 0$, $E\xi_k^2 = \sigma_\xi^2$. It can be written as the IAR(∞) equation in (3.26) with b_j generated by $B(z) = 1 - (1 - z)^d = \sum_{j=1}^{\infty} b_j z^j$. The transfer function $A(x) = (1 - B(\mathrm{e}^{-\mathrm{i}x}))^{-1}$ satisfies $|A(x)| = |1 - \mathrm{e}^{-\mathrm{i}x}|^{-2d} \sim |x|^{-2d}$, as $x \to 0$, and is integrable for $d \in (0, 1/2)$. The coefficients b_j and g_j of the generating functions $B(z)$ and $G(z) = (1 - B(z))^{-1} = (1 - z)^{-d}$ are given by

$$b_j = -\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}, \quad g_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}, \quad j \geq 1, \quad g_0 = 1. \tag{3.32}$$

They have properties $b_j > 0$, $g_j > 0$, $\theta = \sum_{j=1}^{\infty} b_j = 1$, and

$$b_j \sim -j^{-d-1}/\Gamma(-d), \quad g_j \sim j^{d-1}/\Gamma(-d), \qquad j \to \infty, \tag{3.33}$$

so that $\|g\| < \infty$. Relations (3.33) imply that the covariance

$$\gamma_k = \mathrm{cov}(x_0, x_k) = \sigma_\xi^2 \sum_{j=0}^{\infty} g_j g_{k+j}$$

decays hyperbolically, that is,

$$\gamma_k \sim c_\gamma k^{-1+2d}, \qquad c_\gamma = \frac{\sigma_\xi^2 \Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}, \tag{3.34}$$

and the spectral density is singular at the origin:

$$f(x) = (\sigma_\xi^2/2\pi)|1 - \mathrm{e}^{\mathrm{i}x}|^{-2d} \sim c_f |x|^{-2d}, \; c_f = \sigma_\xi^2/2\pi.$$

Example 3.2. A nonparametric (depending on an infinite number of parameters) class of IAR processes $x_k = \sum_{j=1}^{\infty} b_j x_{k-j} + \xi_k$ generalizing the previous example is defined by equation (3.26) with uncorrelated noise $\{\xi_k, k \in \mathbb{Z}\}$ and coefficients b_j generated by the operator

$$B(L) = \big(1 - (1-L)^d\big)P(L) = \sum_{j=1}^{\infty} b_j L^j, \quad 0 < d < 1/2. \tag{3.35}$$

Here, $P(z) = \sum_{j=0}^{\infty} p_j z^j$ is a generating function with coefficients satisfying

$$p_j \geq 0, \quad p_0 > 0, \quad \sum_{j=0}^{\infty} p_j = 1 \quad \text{and} \quad \sum_{j=1}^{\infty} j p_j < \infty. \tag{3.36}$$

Then $b_j = \sum_{k=0}^{j-1} p_k b^0_{j-k}$, where b^0_j are the coefficients of the expansion $1-(1-z)^d = \sum_{j=1}^{\infty} b^0_j z^j$, see (3.32). Hence, the b_j in (3.35) are nonnegative and sum up to 1.

Let us show that $|A(x)|^2 = |1-B(\mathrm{e}^{\mathrm{i}x})|^{-2}$ is integrable. Since $b_1 = p_0 b^0_1 > 0$, by Lemma 3.1 (ii) $|A(x)|$ is bounded on $[\epsilon, \pi]$ for any $\epsilon > 0$. Therefore, it suffices to show that $|A(x)|^2$ is integrable at $x = 0$. To this end, rewrite $1 - B(\mathrm{e}^{\mathrm{i}x}) = 1 - \big(1 - (1-\mathrm{e}^{\mathrm{i}x})^d\big) P(\mathrm{e}^{\mathrm{i}x}) = (1-\mathrm{e}^{\mathrm{i}x})^d h(x)$, where

$$h(x) = P(\mathrm{e}^{\mathrm{i}x}) - \big(P(\mathrm{e}^{\mathrm{i}x}) - 1\big)(1-\mathrm{e}^{\mathrm{i}x})^{-d}. \tag{3.37}$$

From (3.36) we have $|P(\mathrm{e}^{\mathrm{i}x}) - 1| = \sum_{j=1}^{\infty} |\mathrm{e}^{\mathrm{i}jx} - 1| p_j \leq |x| \sum_{j=1}^{\infty} j p_j = O(|x|) = o(|(1-\mathrm{e}^{\mathrm{i}x})^d|)$ and, therefore, $\lim_{x\to 0} h(x) = h(0) = P(1) = 1$. Hence, $|A(x)|^2 \sim |x|^{-2d}$, $x \to 0$, proving the integrability of $|A(x)|^2$ for $d \in (0, 1/2)$. The corresponding stationary solution $\{x_k, k \in \mathbb{Z}\}$ of (3.26) with uncorrelated noise $\{\xi_k\}$ has spectral density

$$f(x) = (\sigma^2_\xi/2\pi)|1 - B(\mathrm{e}^{-\mathrm{i}x})|^{-2} = (\sigma^2_\xi/2\pi)|1-\mathrm{e}^{-\mathrm{i}x}|^{-2d}|h(x)|^{-2}, \quad x \in \Pi, \tag{3.38}$$

with h defined at (3.37). It satisfies $f(x) \sim (\sigma^2_\xi/2\pi)|x|^{-2d}, x \to 0$, and is a continuous bounded function on intervals $[\epsilon, \pi], \epsilon > 0$. Moreover, using (3.38), (3.37), (3.36) and Lemma 2.3.1 of Giraitis *et al.* [36], one can show that the asymptotics of the covariance function is $\mathrm{cov}(x_0, x_k) \sim c_\gamma k^{-1+2d}, k \to \infty$, with c_γ given in (3.34) is the same as for ARFIMA$(0, d, 0)$ model. Hence, the p_j or $P(L)$ in (3.35) essentially affects short memory dynamics and do not affect the long-run behavior of the corresponding

IAR process.

Example 3.3. (The IAR(q, d) model). We introduce the parametric class IAR(q, d) consisting of IAR(∞) processes (3.26) with $B(L)$ as in (3.35) and $P(L)$ a polynomial of degree q satisfying (3.36). It is convenient to parameterize such polynomials as

$$P(z) = \frac{1 + r_1 z + \cdots + r_q z^q}{1 + r_1 + \cdots + r_q}, \quad r_1 \geq 0, \ldots, r_q \geq 0.$$

Thus, $p_i = r_i/(1 + \cdots + r_q)$, $1 \leq i \leq q$, $p_0 = 1/(1 + \cdots + r_q)$, satisfy (3.36) so that IAR(q, d) is a particular case and shares the same long memory properties as IAR in Example 3.2. Note that the IAR($0, d$) model coincides with ARFIMA($0, d, 0$). Apart from this case, it seems that the IAR(q, d) models are different from the ARFIMA(p, d, q) models. For example, the model $(1 - B(L))x_k = \xi_k$ with $B(z) = (1 - (1 - z)^d)(1 + rz)/(1 + r)$ with $P(z) = (1 + rz)/(1 + r)$ generates a different covariance structure than the ARFIMA($1, d, 0$) model $(1 - L)^d(1 + rL)x_k = \xi_k$.

3.4 Proofs of Theorem 3.1 and Corollaries 3.1-3.4

The following proposition used to prove Theorem 3.1 establishes the relation between solutions τ_k of ARCH(∞) equation (3.1), and (Y_k, z_k) of bilinear equations (3.15) and (3.17), with ε_k and ζ_k related by $\varepsilon_k = \sigma\zeta_k + 1$, and $\omega = \mu(1 - \theta)$. For Y_k in (3.15), we define "noise" as $z_k = \zeta_k(\mu\sigma + \sigma \sum_{j=1}^{\infty} b_j Y_{k-j})$. For z_k in (3.17), the volatility process v_k is defined in (3.18).

Proposition 3.1. *Let* $0 < \mu < \infty$ *and* $\theta \in (0, 1]$.

(i) *If* $\{\tau_k, k \in \mathbb{Z}\}$ *is a causal* L^2*-solution of (3.1), then* $\{Y_k = \tau_k - \mu, k \in \mathbb{Z}\}$ *is a causal* L^2*-solution of (3.15) such that* $Y_k \geq -\mu$.

(ii) *If* $\{Y_k, k \in \mathbb{Z}\}$ *is a causal* L^2*-solution of (3.15) such that* $Y_k \geq -\mu$, *then* $\{\tau_k = Y_k + \mu, k \in \mathbb{Z}\}$ *is a causal* L^2*-solution of equation (3.1).*

(iii) $\{Y_k, k \in \mathbb{Z}\}$ *is a causal* L^2*-solution of bilinear equation (3.15) if and only if* $\{Y_k, z_k, k \in \mathbb{Z}\}$ *is a causal* L^2*-solution of equation (3.17). Moreover,* $\{Y_k \geq -\mu\}$ *is equivalent to* $\{v_k \geq 0\}$ *with* v_k *as in (3.18).*

Proof. The equivalence of (i) and (ii) is immediate. We only need to prove (iii). Let $\{Y_k, k \in \mathbb{Z}\}$ be a causal L^2-solution of (3.15). Set $z_k = \zeta_k(\mu\sigma + \sigma \sum_{j=1}^{\infty} b_j Y_{k-j})$ and denote $v_k = \mu\sigma + \sum_{j=1}^{\infty} h_j z_{k-j}$. Let us prove that $\{Y_k, z_k, k \in \mathbb{Z}\}$ is a causal L^2-solution of (3.17). This follows from (3.15) and equality

$$v_k = \mu\sigma + \sigma \sum_{j=1}^{\infty} b_j Y_{k-j}, \tag{3.39}$$

which is verified below. From the definition of z_k and (3.15) it follows that Y_k satisfy the IAR equation $Y_k - \sum_{j=1}^{\infty} b_j Y_{t-j} = z_k$, where $\{z_k, k \in \mathbb{Z}\}$ is a causal uncorrelated process with finite variance. Therefore, by Theorem 3.2 we have $Y_k = \sum_{j=0}^{\infty} g_j z_{k-j}$, which implies that $\sigma \sum_{j=1}^{\infty} b_j Y_{k-j} = \sigma \sum_{j=1}^{\infty} b_j \sum_{i=0}^{\infty} g_i z_{k-j-i} = \sum_{j=1}^{\infty} h_j z_{k-j}$ in view of the definition of h_j in (3.16), proving (3.39) and the fact that $\{Y_k, z_k\}$ is a causal L^2-solution of (3.17). Moreover, $Y_{k-j} \geq -\mu$, $k \in \mathbb{Z}$, and (3.39) imply $v_k \geq \mu\sigma + \sigma(\sum_{j=1}^{\infty} b_j)(-\mu) = \mu\sigma(1-\theta) \geq 0$, $k \in \mathbb{Z}$.

Conversely, assume that $\{Y_k, z_k, k \in \mathbb{Z}\}$ is a causal L^2-solution of (3.17). Then the claim that $\{Y_k, k \in \mathbb{Z}\}$ is a causal L^2-solution of (3.15) follows from (3.39), which, in turn, follows from Theorem 3.2 using exactly the same argument as above. Finally, from $v_k \geq 0$, (3.39), (3.15) and $\zeta_k \geq -1/\sigma$, we obtain

$$\begin{aligned} Y_k &= \sum_{j=0}^{\infty} b_j Y_{k-j} + \zeta_k v_k \ \geq \ \sum_{j=0}^{\infty} b_j Y_{k-j} - (1/\sigma) v_k \\ &= \sum_{j=0}^{\infty} b_j Y_{k-j} - (1/\sigma)(\mu\sigma + \sigma \sum_{j=0}^{\infty} b_j Y_{k-j}) \ = \ -\mu, \end{aligned}$$

proving part (iii) and the proposition. □

Proof of Theorem 3.1. (a) The equivalence of (3.22) and (3.23) follows from the equivalence of $\|g\| < \infty$ and $\|A\| < \infty$, see Lemma 3.2, and Parseval's identity $\|g\| = 2\pi\|A\|$. Let us prove the necessity of condition (3.22), or $\|h\| < 1$, for the existence of a stationary solution. Assume that $\{\tau_k, k \in \mathbb{Z}\}$ is an L^2-solution of ARCH equation (3.1). Then, by Proposition 3.1 (i), the last fact implies that for $\mu > 0$, $\{Y_k = \tau_k - \mu, z_k = \zeta_k(\mu\sigma + \sigma \sum_{j=1}^{\infty} b_j Y_{k-j}), k \in \mathbb{Z}\}$ is an L^2-solution of bilinear equation (3.17). Consequently, $\sigma_z^2 = E z_k^2 = E\big(\mu\sigma + \sum_{j=1}^{\infty} h_j z_{k-j}\big)^2 = (\mu\sigma)^2 + (\sum_{j=1}^{\infty} h_j^2)\sigma_z^2 = (\mu\sigma)^2 + \|h\|^2\sigma_z^2$, yielding $\|h\|^2 < 1$, or (3.22), since $\|h\|^2 = \sigma^2(\|g\|^2 - 1)$.

Conversely, let us show that $\|h\| < 1$ implies the existence of the L^2-solution $\{\tau_k, k \in \mathbb{Z}\}$ of (3.1) with $E\tau_k = \mu$ given by $\tau_k = Y_k + \mu$ and Y_k defined in (3.21), (3.19). As shown in (3.20), $\|h\| < 1$ guarantees that $\{Y_k\}$ is an L^2-solution of (3.15). Therefore, by Proposition 3.1 (ii), it suffices to prove that

$$Y_k \geq -\mu. \tag{3.40}$$

To show (3.40), we approximate Y_k by

$$Y_{k,p} = (\mu\sigma) \sum_{m=1}^{\infty} \left(\sum_{p < s_m < \cdots < s_1 \leq k} g_{k-s_1} h_{s_1-s_2} \cdots h_{s_{m-1}-s_m} \zeta_{s_1} \cdots \zeta_{s_m} \right),$$

where $-p \geq 1$ is a large integer. Observe that for $k > p$ the $Y_{k,p}$ satisfy equation (3.15), viz.,

$$Y_{k,p} \;=\; \zeta_k \left(\mu\sigma + \sigma \sum_{j=1}^{\infty} b_j Y_{k-j,p} \right) + \sum_{j=1}^{\infty} b_j Y_{k-j,p}, \quad \text{for } k > p, \tag{3.41}$$

while $Y_{k,p} = 0$ for $k \leq p$. Moreover, by orthogonality of Volterra series, $E(Y_k - Y_{k,p})^2 = (\mu\sigma)^2 \sum_{m=1}^{\infty} J_{k,p}^{(m)}$, where

$$J_{k,p}^{(m)} = \sum_{s_m < \cdots < s_1 \leq k,\, s_m \leq p} g_{k-s_1}^2 h_{s_1-s_2}^2 \cdots h_{s_{m-1}-s_m}^2 .$$

Notice that $J_{k,p}^{(m)} \leq \|g\|^2 \|h\|^{2(m-1)}$, where $\|h\| < 1$. Hence, $\sum_{m=1}^{\infty} J_{k,p}^{(m)}$ is dominated by a converging series. Moreover, for each $m \geq 1$, $J_{k,p}^{(m)} \to 0$ as $p \to -\infty$. Hence, $\lim_{p\to-\infty} E(Y_k - Y_{k,p})^2 = 0$ for any $k \in \mathbb{Z}$ by the dominated convergence theorem. Therefore, (3.40) follows if we show that for any $p \in \mathbb{Z}$,

$$Y_{k,p} \geq -\mu, \qquad k \in \mathbb{Z}. \tag{3.42}$$

To prove (3.42), we use induction on k. Clearly, (3.42) holds for $k \leq p$ because by definition $Y_{k,p} = 0 > -\mu$ for $k \leq p$. Also, (3.42) holds for $k = p+1$, since $Y_{p+1,p} = (\mu\sigma)\zeta_{p+1} \geq -\mu$ because $(\mu\sigma)\zeta_j = (\mu\sigma)(\varepsilon_j - 1)/\sigma \geq -\mu$, for $j \in \mathbb{Z}$. Let $k > p+1$. Assume by induction that $Y_{s,p} \geq -\mu$ for all $s < k$. Then, by (3.41) and the inductive assumption,

$$\begin{aligned} Y_{k,p} &= \zeta_k(\mu\sigma) + (\zeta_k\sigma + 1)\left(\sum_{j=1}^{\infty} b_j Y_{k-j,p}\right) \geq \\ &\geq \zeta_k(\mu\sigma) + (\zeta_k\sigma + 1)\left(\sum_{j=1}^{\infty} b_j\right)(-\mu) = -\mu. \end{aligned}$$

This proves the induction step $k-1 \to k$ as well as (3.42) and (3.40), thereby proving part (a) of the theorem.

(b) Claim (i) is shown in Giraitis *et al.* [30], Theorem 3.1. Let us prove (ii). By part (a), it suffices to prove the uniqueness of the solution $\{\tau_k, k \in \mathbb{Z}\}$. Let $\{\tau'_k, k \in \mathbb{Z}\}, \{\tau''_k, k \in \mathbb{Z}\}$ be two causal L^2-solutions of (3.1) with $E\tau'_k = E\tau''_k$. Then $\tau'_k - \tau''_k = Y_k$ is a causal L^2-solution of $Y_k = \sum_{j=1}^\infty b_j Y_{k-j} + z_k$, where $z_k = \zeta_k \sigma \sum_{j=1}^\infty b_j Y_{k-j}$. By causality, the stationary process $Y_k = f(\varepsilon_k, \varepsilon_{k-1}, \dots)$ is a function of lagged i.i.d. variables. Hence, $\{Y_k\}$ is a regular process with $EY_k^2 < \infty$, having spectral density, see Ibragimov and Linnik [48], Theorem 17.1.2. Moreover, $z_k = \zeta_k \sum_{j=1}^\infty h_j z_{k-j}$, see (3.17) (b), where $\{z_k, k \in \mathbb{Z}\}$ is covariance stationary white noise and $\sum_{j=1}^\infty h_j^2 = \|h\|^2 < 1$, $E\zeta_k^2 = 1$. Then $Ez_k^2 = \sum_{j=1}^\infty h_j^2 E z_{k-j}^2 = \|h\|^2 E z_k^2$ implies $Ez_k^2 = 0$ and hence $z_k = 0$. Therefore, $\{Y_k, k \in \mathbb{Z}\}$ has spectral density and is a stationary solution of the homogeneous equation $Y_k - \sum_{j=1}^\infty b_j Y_{k-j} = 0$. As shown in the proof of Theorem 3.2 (ii) below, such an equation has a unique solution $Y_k \equiv 0$, proving the uniqueness of $\{\tau_k, k \in \mathbb{Z}\}$. Theorem 3.1 is proved. □

Proof of Corollary 3.1. All claims with the exception of (iii) follow from Theorem 3.1, and claim (iii) follows from Theorem 3.2 (iii). □

Proof of Corollary 3.2. Note $\sigma^2 = E\varepsilon_0^2 - 1$. We have $\|A\|^2 = \int_\Pi |1 - \mathrm{e}^{\mathrm{i}x}|^{-2d} \mathrm{d}x = 2\pi\Gamma(1-2d)/\Gamma^2(1-d)$ yielding the equivalence of (3.23) and (3.8). The remaining claims follow from Corollary 3.1 and fact (3.34) in Example 3.1. □

Proof of Corollary 3.3. All statements with the exception of the last claim follow from Theorem 3.1. To show it, note that $g_j \geq 0$ in (3.27) satisfy $\sum_{j=0}^\infty g_j \leq \sum_{m=0}^\infty \theta^m < \infty$ since $\theta < 1$. □

Proof of Corollary 3.4. It suffices to only show part (ii). Using the fact that by (3.21) $Y_k = \tau_k - E\tau_k = \sum_{j=0}^{\infty} g_j z_{k-j}$ is a moving average in stationary ergodic martingale differences $\{z_s\}$ of (3.19) with coefficients g_j given in (3.32) and satisfying (3.33), the convergence in (ii) follows from Theorem 3.1 in Abadir *et al.* [1] or Theorem 6.2 in Giraitis and Surgailis [30]. □

Proof of Corollary 3.5. Let $\theta = 1$. Then the LM(d)-ARCH model in (3.6) coincides with the FIGARCH model in (3.5) with $\omega = 0$ and the statement follows from Corollary 3.2. Next, let $\theta < 1$. Then (3.6) can be written as the ARCH(∞) equation in (3.1) with $\omega = \mu(1-\theta) > 0$. According to Corollary 3.3, the squared process $\{r_k^2 = \tau_k\}$ has short memory and summable autocovariance $\sum_{k=0}^{\infty} \operatorname{cov}(r_0^2, r_k^2) < \infty$ which contradicts (3.7).
□

3.5 Proofs of Theorems 3.2 and 3.3

The proof of Theorem 3.1 uses auxiliary Lemmas 3.1 and 3.2. The proofs of these lemmas are provided at the end of this section. Denote $J_b = \{j \geq 1 : b_j > 0\}$, and assume J_b has at least two elements. Denote by $\operatorname{gcd}(J_b)$ the greatest common divisor of $j \in J_b$. For example, if $b_1 > 0$, then $\operatorname{gcd}(J_b) = 1$, and if $b_{2j} > 0$, $b_{2j-1} = 0$, $j = 1, 2, \ldots$, then $\operatorname{gcd}(J_b) = 2$.

Lemma 3.1. *Let* $\theta = \sum_{j=1}^{\infty} b_j = 1$.

(i) *The function* $1 - B(\mathrm{e}^{\mathrm{i}x})$, $x \in \Pi$, *has only finite number of zeroes on* Π, *including* $x = 0$.

(ii) *The point* $x = 0$ *is the unique zero of* $1 - B(\mathrm{e}^{\mathrm{i}x})$ *if and only if* $\operatorname{gcd}(J_b) = 1$.

(iii) *If* $b_k = O(k^{-\gamma})$, $k \to \infty$, *for some* $\gamma \geq 3/2$, *then* $\|A\| = \infty$.

Lemma 3.2. *Let $\theta \le 1$.*

(i) *If $\|g\| < \infty$ then $\|A\| < \infty$ and (3.29) hold.*

(ii) *If $\|A\| < \infty$ then $\|g\| < \infty$.*

Proof of Theorem 3.2. All statements in (iv) follow from Lemma 3.2.

(i) If $\|g\| < \infty$, then by (iv), $\|A\| < \infty$ and (3.29) holds. Evidently, this implies that (3.28) is a stationary solution of (3.26).

Conversely, if a stationary solution $\{x_k, k \in \mathbb{Z}\}$ of (3.26) exists, it suffices to prove that $\|A\| < \infty$, which by (iv) implies $\|g\| < \infty$. Let $x_k = \int_\Pi \mathrm{e}^{\mathrm{i}ky} Z_x(\mathrm{d}y)$ be the spectral representation of $\{x_k, k \in \mathbb{Z}\}$ and $F_x(\mathrm{d}y) = E|Z_x(\mathrm{d}y)|^2$ be its spectral measure (we do not assume *a priori* that $\{x_k, k \in \mathbb{Z}\}$ has spectral density). Denote by $\xi_k = \int_\Pi \mathrm{e}^{\mathrm{i}ky} Z_\xi(\mathrm{d}y)$ the spectral representation of the noise $\{\xi_k, k \in \mathbb{Z}\}$ and by $F_\xi(\mathrm{d}y) = E|Z_\xi(\mathrm{d}y)|^2 = (\sigma_\xi^2/2\pi)\mathrm{d}y$ its spectral measure. Since the series $B(\mathrm{e}^{\mathrm{i}y}) = \sum_{j=1}^\infty b_j \mathrm{e}^{\mathrm{i}jy}$ converges uniformly in Π to a bounded function, $x_k - \sum_{j=1}^\infty b_j x_{k-j} = \int_\Pi (1 - B(\mathrm{e}^{\mathrm{i}y})) Z_x(\mathrm{d}y) = \xi_k = \int_\Pi \mathrm{e}^{\mathrm{i}ky} Z_\xi(\mathrm{d}y)$, leading to

$$|1 - B(\mathrm{e}^{\mathrm{i}y})|^2 F_x(\mathrm{d}y) = F_\xi(\mathrm{d}y) = (\sigma_\xi^2/2\pi)\mathrm{d}y, \quad y \in \Pi. \tag{3.43}$$

By Lemma 3.1 (i), $1 - B(\mathrm{e}^{-\mathrm{i}y})$ has a finite number of zeros $y_1, \dots, y_m \in \Pi$. Since F_x is nondecreasing, (3.43) implies that $F_x(\mathrm{d}y)$ coincides with $f(y)\mathrm{d}y$, $f(y) = (\sigma_\xi^2/2\pi)|A(y)|^2$, except for possible jumps at points $y_1, \dots,$ y_m, i.e. $F_x(\mathrm{d}y) = f(y)\mathrm{d}y + \sum_{i=1}^m c_i \delta_{y_i}$, where $c_i \ge 0$ are some nonnegative constants. Therefore,

$$\infty > E x_k^2 = \int_\Pi F_x(\mathrm{d}y) \ge \int_\Pi f(y)\mathrm{d}y = (\sigma_\xi^2/2\pi) \int_\Pi |A(y)|^2 \mathrm{d}y,$$

proving $\|A\| < \infty$.

(ii) Since $\{\tilde{x}_k, k \in \mathbb{Z}\}$ in (3.28) is a zero mean L^2-solution of equation (3.26), see the proof of (i) above, it remains to show the uniqueness of solution $x_k = \mu + \tilde{x}_k$ of (3.26) with the stated properties. Let $\{x'_k, k \in \mathbb{Z}\}, \{x''_k, k \in \mathbb{Z}\}$, be two stationary L^2-solutions of (3.26) with $Ex'_k = Ex''_k$ and let $y_k = x'_k - x''_k$. Moreover, by the assumption in (ii), y_k has the form $y_k = \sum_{j\in\mathbb{Z}} c_j \xi_{k-j}$ with $\sum_{j\in\mathbb{Z}} c_j^2 < \infty$. The above facts imply that $\{y_k, k \in \mathbb{Z}\}$ is a L^2-solution of the homogeneous equation $y_k - \sum_{j=1}^{\infty} b_j y_{k-j} = 0$ and a stationary process with absolutely continuous spectral measure $F_y(\mathrm{d}x) = f_y(x)\mathrm{d}x$, $f_y(x) = (\sigma_\xi^2/2\pi)|\sum_{j\in\mathbb{Z}} c_j \mathrm{e}^{\mathrm{i}jx}|^2$ and a spectral representation $y_k = \int_\Pi \mathrm{e}^{\mathrm{i}kx} Z_y(\mathrm{d}x)$. Since $\sum_{j=1}^{\infty} \mathrm{e}^{\mathrm{i}jx} b_j$ converges uniformly on Π, hence also in $L^2(F_y)$, it follows that $y_k - \sum_{j=1}^{\infty} b_j y_{k-j} = \int_\Pi (1 - B(\mathrm{e}^{\mathrm{i}x}))Z_y(\mathrm{d}x) = 0$ and $\int_\Pi |1 - B(\mathrm{e}^{\mathrm{i}x})|^2 F_y(\mathrm{d}x) = 0$. Together with Lemma 3.1 (i), this implies that $f_y(x) = 0$ a.e. on Π and hence $F_y = 0$ and $y_k = 0$, proving part (ii).

(iii) As noted above, solution $\tilde{x}_k$ in (3.28) has spectral density $f(x) = (\sigma_\xi^2/2\pi)|1 - B(\mathrm{e}^{\mathrm{i}x})|^{-2}$. Relation $\lim_{x\to 0} f(x) = \infty$ follows from $B(1) = 1$, continuity of $B(\mathrm{e}^{-\mathrm{i}x})$, and the fact $|B(x)| < 1$ for $0 < x < x_0$ for some $x_0 > 0$ which holds by Lemma 3.1 (i). The divergence $\sum_{k\in\mathbb{Z}} |\mathrm{cov}(x_0, x_k)| = \infty$ is immediate from the previous fact. Finally, the first claim in (3.31) is a consequence of moving average representation (3.28) and positivity of g_j. Theorem 3.2 is proved. □

Proof of Theorem 3.3. The proof follows using the same arguments as in the proof of Theorem 3.2. □

Proof of Lemma 3.1. (i) First observe that $B(\mathrm{e}^{\mathrm{i}x}) = \theta = 1$ holds for $x = 0$. Suppose that $x \in (0, 2\pi)$ is such that $1 = B(\mathrm{e}^{\mathrm{i}x})$. Then $B(\mathrm{e}^{\mathrm{i}x})$ is a real number: $B(\mathrm{e}^{\mathrm{i}x}) = \sum_{j=1}^{\infty} b_j \cos(jx)$ and then $1 = \sum_{j=1}^{\infty} b_j \cos(jx) \le$

$\sum_{j=1}^{\infty} b_j = 1$ is possible if and only if $1 = \cos(jx) = \mathrm{e}^{\mathrm{i}jx}$ for all $j \in J_b$, or

$$x/2\pi \in \bigcap_{j \in J_b} I_j, \quad \text{where} \quad I_j = \left\{ \frac{1}{j}, \frac{2}{j}, \dots, \frac{j-1}{j} \right\}. \tag{3.44}$$

Clearly, since each I_j is a finite set, the intersection in (3.44) is a finite set as well, proving (i).

(ii) Let $\gcd(J_b) = 1$. Then $\gcd(j_1, j_2) = 1$ for $j_1, j_2 \in J_b, j_1 \neq j_2$. It suffices to show that $I_{j_1} \cap I_{j_2} = \emptyset$. Indeed, assume *ad absurdum* that $I_{j_1} \cap I_{j_2} \neq \emptyset$, then $k_2 = k_1 j_2 / j_1$ for some integers $1 \le k_1 < j_1$, $1 \le k_2 < j_2$, by definition of I_j in (3.44). Since j_1 and j_2 are coprimes, this means that j_1 is a divisor of k_1, or $k_1 \in \{j_1, 2j_1, \dots\}$, which contradicts $k_1 < j_1$.

Let $p = \gcd(J_b) \ge 2$. Then for any $j \in J_b$, $j = j'p$ with $1 \le j' < j$. Thus, $j/p \in \{1, 2, \dots, j-1\}$, implying $1/p \in I_j$ for all $j \in J_b$ and $1/p \in \bigcap_{j \in J_b} I_j$. Particularly, $x = 2\pi/p \neq 0$ is a zero of $1 - B(\mathrm{e}^{\mathrm{i}x})$.

(iii) It suffices to show $|1 - B(\mathrm{e}^{\mathrm{i}x})| \le C|x|^{1/2}$ as this implies $\int_\Pi |1 - B(\mathrm{e}^{\mathrm{i}x})|^{-2} \mathrm{d}x \ge C^{-1} \int_\Pi \mathrm{d}x/|x| = \infty$. We have $|1 - \mathrm{e}^{\mathrm{i}jx}| \le \min(j|x|, 2)$, $b_j \le Cj^{-\gamma} \le Cj^{-3/2}$ and thus

$$\begin{aligned} \left|1 - B(\mathrm{e}^{\mathrm{i}x})\right| &= \left|\sum_{j=1}^{\infty} b_j (1 - \mathrm{e}^{\mathrm{i}jx})\right| \le C|x| \sum_{1 \le j < 1/|x|} j^{1/2} + \\ &+ C \sum_{j \ge 1/|x|} j^{-3/2} \le C|x|^{1/2}. \end{aligned}$$

This proves (iii) and the lemma, too. □

Proof of Lemma 3.2. (i) Suppose that $\|g\| < \infty$. Set $k_r(x) = \sum_{j=0}^{\infty} g_j r^j \mathrm{e}^{\mathrm{i}jx}$,

$0 < r < 1$. Then $\{k_r\}$ is a Cauchy sequence in $L_2(\Pi)$:

$$\|k_r - k_{r'}\|^2 = \int_\Pi \Big| \sum_{j=0}^\infty g_j(r^j - r'^j)\mathrm{e}^{\mathrm{i}jx}\Big|^2 \mathrm{d}x = 2\pi \sum_{j=0}^\infty g_j^2 |r^j - r'^j|^2 \to 0,$$

as $r, r' \uparrow 1$. Moreover, $k_r(x) \to A(x) = (1 - B(\mathrm{e}^{\mathrm{i}x}))^{-1}$ a.e. in Π as $r \uparrow 1$, since $k_r(x) = G(r\mathrm{e}^{\mathrm{i}x}) = (1 - B(r\mathrm{e}^{\mathrm{i}x}))^{-1}$ for $0 < r < 1$ and $1 - B(\mathrm{e}^{\mathrm{i}x}) \neq 0$ a.e. in Π (see Lemma 3.1 (i)). Therefore, $\|k_r - A\| \to 0$ as $r \uparrow 1$ and $\|A\| < \infty$, see Rudin ([65], Thorem 3.12). Since $k_1 \in L_2(\Pi)$ and $\|k_r - k_1\|^2 \to 0$ as $r \uparrow 1$, then $A = k_1$ in $L_2(\Pi)$ which proves (3.29).

(ii) Let $\|A\| < \infty$. Then functions $h_k(x) = \mathrm{e}^{-\mathrm{i}kx}/(1 - B(\mathrm{e}^{\mathrm{i}kx})) = \mathrm{e}^{-\mathrm{i}kx}A(x)$, $x \in \Pi$, $k \in \mathbb{Z}$ belong to the Hilbert space $L^2(\Pi)$ with the norm

$$\|h\| = \Big(\int_\Pi |h(x)|^2 \mathrm{d}x\Big)^{1/2}.$$

So, $\|h_k\| = \|A\| < \infty$. Then $h_k(x) - \sum_{j=1}^\infty b_j h_{k-j}(x) = \mathrm{e}^{\mathrm{i}kx}$, where the series converges in $L^2(\Pi)$. By Lemma 3.1 (iii), $\|A\| < \infty$ implies that $b_j > 0$ for infinite number of j. For a large $p \geq 1$ denote $b'_j = b_j I(j \leq p)$ and $B'(\mathrm{e}^{\mathrm{i}x}) = \sum_{j=1}^\infty b'_j \mathrm{e}^{\mathrm{i}jx}$. Then

$$h_k(x) - \sum_{j=1}^p b_j h_{k-j}(x) = \mathrm{e}^{\mathrm{i}kx} + u_k(x), \quad \text{where} \quad u_k(x) = \sum_{j=p+1}^\infty b_j h_{k-j}(x).$$

Since $\sum_{j=1}^p b_j h_{k-j}(x) = h_k(x)B'(\mathrm{e}^{\mathrm{i}x})$ and $\sum_{j=1}^p b_j = \sum_{j=1}^\infty b'_j < 1$, we obtain

$$h_k(x) = \frac{\mathrm{e}^{\mathrm{i}kx} + u_k(x)}{1 - B'(\mathrm{e}^{\mathrm{i}x})} = \xi'_k(x) + u'_k(x),$$

where

$$\xi'_k(x) = \frac{\mathrm{e}^{\mathrm{i}kx}}{1 - B'(\mathrm{e}^{\mathrm{i}x})}, \quad u'_k(x) = \frac{u_k(x)}{1 - B'(\mathrm{e}^{\mathrm{i}x})}.$$

We claim that under assumption $\|g\| = \infty$, as $p \to \infty$,

$$\|u'_k\| \le \|h_k\| = ||A|| < \infty \quad \text{and} \quad \|\xi'_k\| \to \infty. \tag{3.45}$$

On the other hand, $\|\xi'_k\| = \|h_k - u'_k\| \le \|h_k\| + \|u'_k\| \le 2\|h_k\| < C < \infty$, which leads to a contradiction, implying $\|g\| < \infty$.

To prove (3.45), note that from the definition of $u'_k(x)$ and

$$1 \le \Big|1 - \sum_{j=1}^{p} b_j \mathrm{e}^{-\mathrm{i}jx}\Big| + \Big|\sum_{j=1}^{p} b_j \mathrm{e}^{-\mathrm{i}jx}\Big| \le \Big|1 - \sum_{j=1}^{p} b_j \mathrm{e}^{-\mathrm{i}jx}\Big| + \sum_{j=1}^{p} b_j$$

we obtain

$$\begin{aligned}
\|u'_k\|^2 &= \int_\Pi \Big|\frac{\sum_{j=p+1}^{\infty} b_j \mathrm{e}^{-\mathrm{i}jx}}{(1 - B(\mathrm{e}^{\mathrm{i}x}))(1 - \sum_{j=1}^{p} b_j \mathrm{e}^{-\mathrm{i}jx})}\Big|^2 \mathrm{d}x \\
&\le \int_\Pi \frac{\mathrm{d}x}{|1 - B(\mathrm{e}^{\mathrm{i}x})|^2} \left(\frac{\sum_{j=p+1}^{\infty} b_j}{1 - \sum_{j=1}^{p} b_j}\right)^2 \le \int_\Pi \frac{\mathrm{d}x}{|1 - B(\mathrm{e}^{\mathrm{i}x})|^2} = \|h_k\|^2,
\end{aligned}$$

proving the first relation in (3.45). The second claim in (3.45) follows from

$$\|\xi'_k\|^2 = \int_\Pi \frac{\mathrm{d}x}{|1 - B'(\mathrm{e}^{\mathrm{i}x})|^2} = \int_\Pi |G'(\mathrm{e}^{\mathrm{i}x})|^2 \mathrm{d}x = 2\pi \sum_{j=0}^{\infty} (g'_j)^2,$$

where g'_j are power coefficients of the analytic function $G'(z) = (1 - B'(z))^{-1} = \sum_{j=0}^{\infty} g'_j z^j$, $|z| < 1$, as given by (3.27) with b_j replaced by b'_j. Note $0 \le g'_j \to g_j$ monotonically as $p \to \infty$ and, therefore, $\sum_{j=0}^{\infty} (g'_j)^2 \to \|g\|^2 = \infty$. This proves part (ii) and completes the proof of the lemma. □

3.6 Simulation study

This section consists of two parts. In the first one, we simulate and compare FIGARCH and ARFIMA$(0, d, 0)$ processes and their estimated autocorrelation functions. The second part is devoted to the investigation of similarities and differences between IAR(p, d, q) and ARFIMA(p, d, q) models.

3.6.1 FIGARCH and ARFIMA(0,d,0) processes

First we describe the simulation procedure. We generate $N = 500$ samples of length $n = 20000$ each, with the first $n_p = 10000$ values being the pre-sample which is not used in the subsequent calculations of autocorrelations. The FIGARCH process was generated using equation

$$\tau_k = \varepsilon_k \left(\sum_{j=1}^{n_p-1+k} b_j \tau_{k-j} + 1 - \sum_{j=1}^{n_p-1+k} b_j \right), \quad -n_p + 1 \leq k \leq n_p, \tag{3.46}$$

and the ARFIMA$(0, d, 0)$ process comes from equation

$$x_k = \zeta_k + \sum_{j=1}^{n_p-1+k} b_j x_{k-j} + \left(1 - \sum_{j=1}^{n_p-1+k} b_j \right), \qquad -n_p + 1 \leq k \leq n_p. \tag{3.47}$$

Initial conditions for both processes are $\tau_k = x_k = 1, k \leq 0$, and coefficients b_j are determined by the generating function $B(z) = \sum_{j=1}^{\infty} b_j z^j = 1 - (1 - z)^d$. We calculate them using the recursive relation

$$\begin{aligned} b_1 &= d, \\ b_j &= b_{j-1} \frac{d - j + 1}{j} (-1)^{2j+1}, \quad j \geq 2. \end{aligned}$$

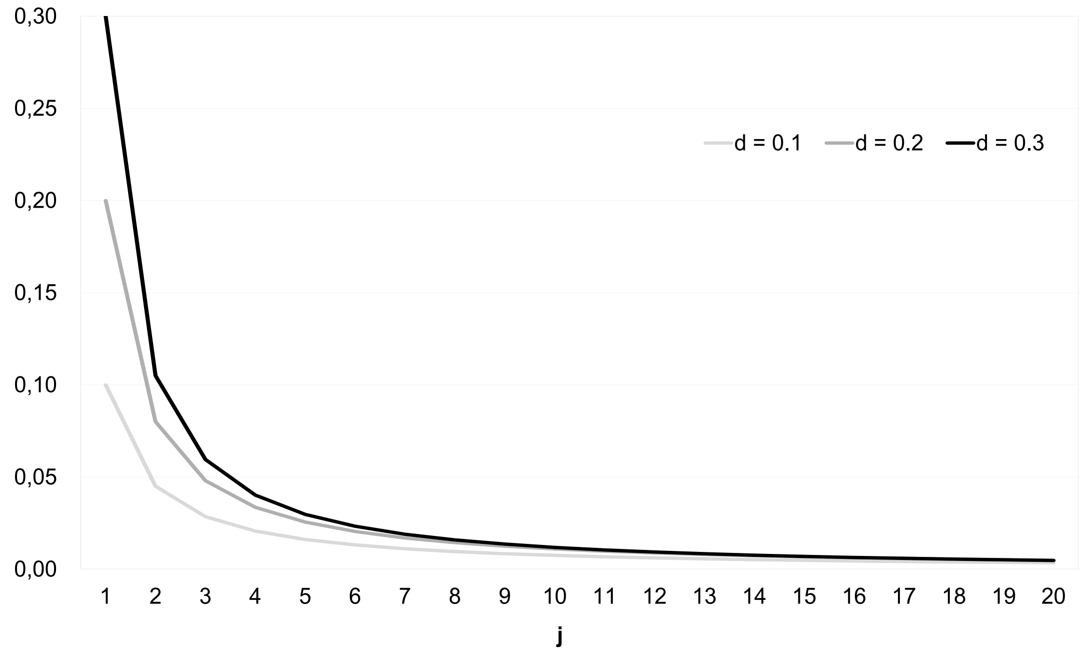

Figure 3.1: Coefficients $b_j, j = 1, \cdots, 20$ for different values of d.

In practice, this way of calculating b_j's seems to be more convenient and accurate compared with the direct formula

$$b_j = -\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)},$$

where we need to deal with very large numbers in the nominator and denominator. The first 20 values of b_j for $d = 0.1, 0.2$ and 0.3 are presented in Figure 3.1. Innovations in (3.47) are i.i.d. standard normal random variables, that is, $\zeta_k \sim N(0,1)$. Then for ε_k in (3.46) we take $\varepsilon_k = \zeta_k^2$, so $\mathrm{E}\varepsilon_k = 1$, $\mathrm{E}\varepsilon_k^2 = 3$ and $\sigma_\varepsilon^2 = 2$. We simulated FIGARCH and ARFIMA$(0, d, 0)$ processes for four different values of $d = 0.01, 0.1, 0.2, 0.3$. In all these cases, the condition for the existence of the stationary FIGARCH solution is satisfied:

$$\mathrm{E}\varepsilon_k^2 = 3 < \frac{\Gamma(1-2d)}{\Gamma(1-2d) - \Gamma^2(1-d)} \approx 4.16, \quad \text{for } d = 0.3.$$

In order to increase comparability between simulated FIGARCH and ARFIMA processes, all the corresponding 500 paths of FIGARCH and ARFIMA$(0, d, 0)$, $\tau_{k,i}, x_{k,i}, k = -n_p + 1, ..., n_p, i = 1, ..., 500$, were simulated using the same set of generated innovations $\zeta^2_{k,i}$ and $\varepsilon_{k,i} = \zeta^2_{k,i}$.

Figure 3.2 exhibits the last 500 values of simulated FIGARCH samples of size 20000 for different values of d. It seems that with higher d, the clusterization of τ_k increases, while the path attains a few higher peaks. Figure 3.3 presents the corresponding samples of the simulated ARFIMA$(0, d, 0)$ process with unit mean. It can be seen that the persistence and peaks in sample paths increase with stronger long memory, that is, with bigger d.

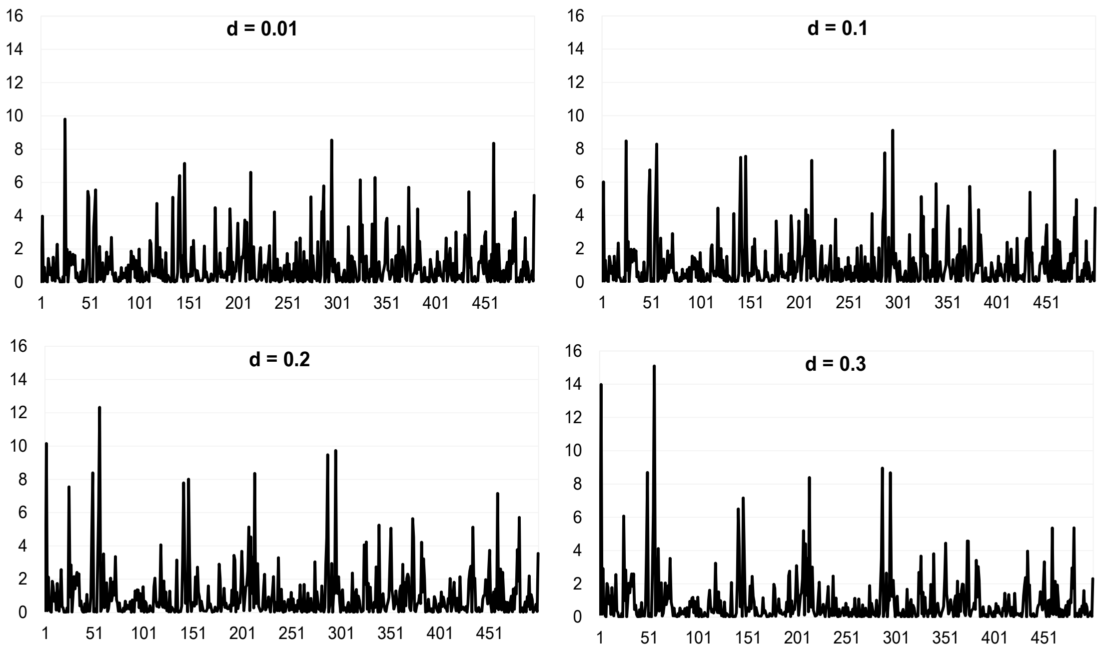

Figure 3.2: The last 500 values of simulated FIGARCH samples of size 20 000 for different values of d ($\mu = 1$).

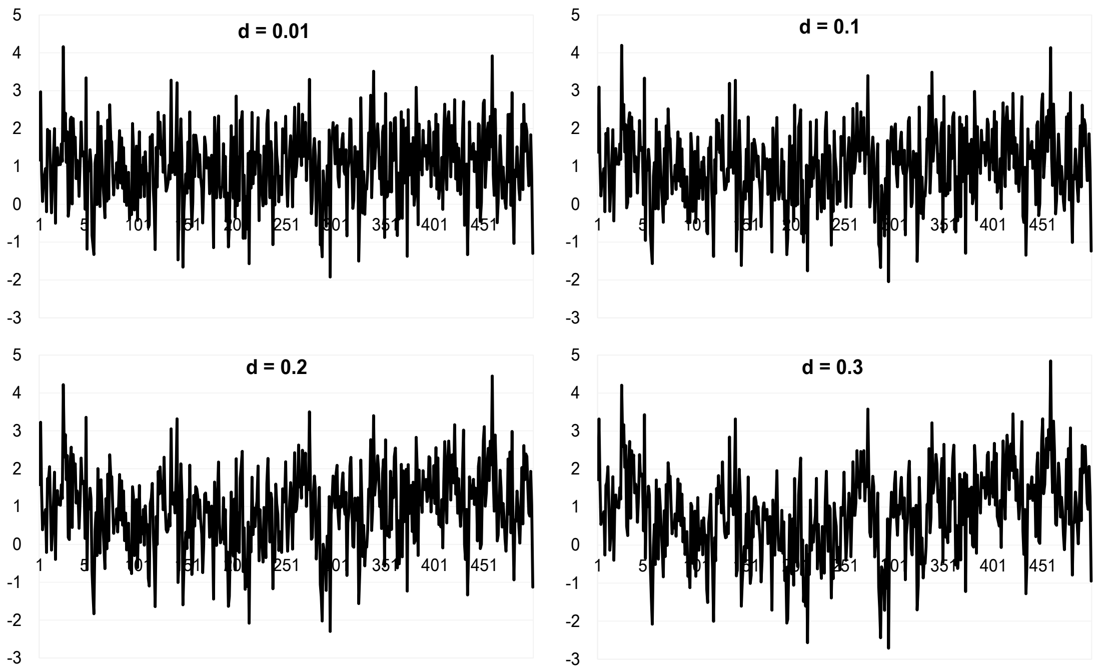

Figure 3.3: The last 500 values of simulated ARFIMA$(0, d, 0)$ samples of size 20 000 for different values of d ($\mu = 1$).

Now we compare the estimated autocorrelation functions (ACFs) of ARFIMA$(0, d, 0)$ and FIGARCH models. The theoretical ACF was estimated using Monte Carlo averaging for 500 independent samples. The first obvious observation from Figure 3.4 is that ACFs markedly increase with d. Another important thing to notice is that for higher d the ARFIMA$(0, d, 0)$ ACF dominates the FIGARCH ACF and the difference increases with d. This fact is rather surprising, as our theoretical results show that these two models share identical asymptotics of the autocorrelation function. Relations (3.25) and (3.34) state that

$$\mathrm{cov}(\tau_0, \tau_k) \sim \mu^2 \sigma_z^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} k^{-1+2d}, \text{ in case of FIGARCH,}$$
$$\mathrm{cov}(x_0, x_k) \sim \sigma_\xi^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} k^{-1+2d}, \text{ in case of ARFIMA}(0, d, 0),$$

where

$$\sigma_z^2 = \sigma^2 / \big(1 + \sigma^2 - \sigma^2(\Gamma(1-2d)/\Gamma^2(1-d))\big), \quad \sigma^2 = \mathrm{Var}(\varepsilon_0),$$

and $\sigma_\xi^2 = \mathrm{Var}(\xi_k)$.

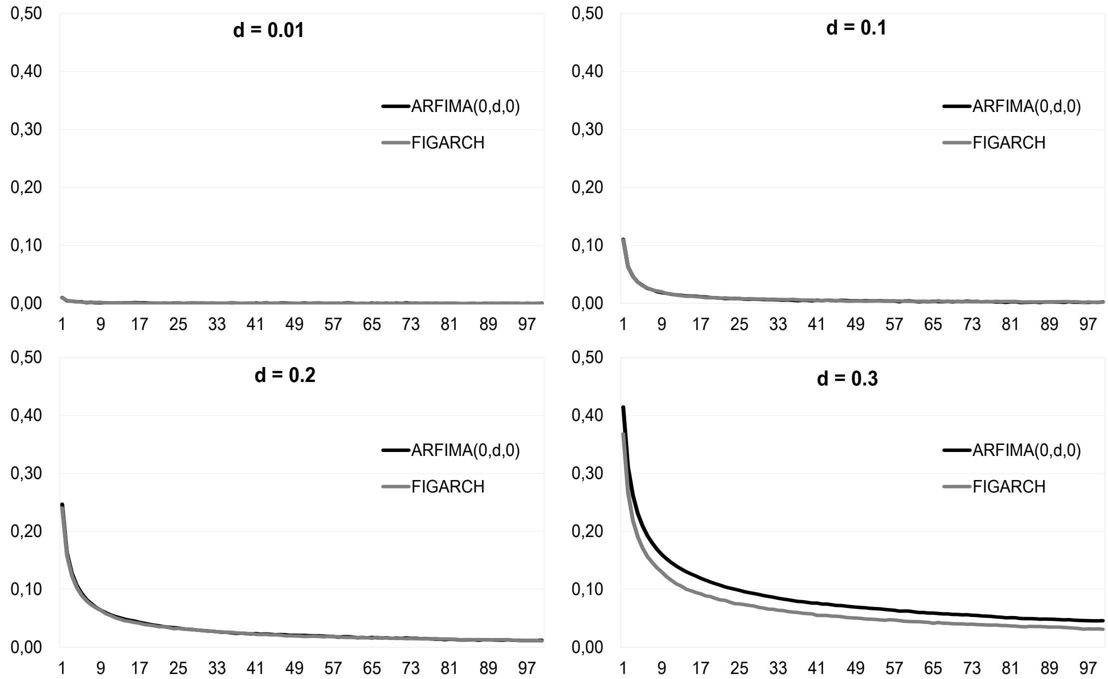

Figure 3.4: Estimated ACF of FIGARCH and ARFIMA$(0, d, 0)$ processes for different values of d (lags $k = 1, ..., 100$).

For spectral densities $f(x)$ and $f^*(x)$ of FIGARCH and ARFIMA$(0, d, 0)$ we have accordingly

$$f(x) = \frac{\sigma_z^2}{2\pi} \left|1 - e^{ix}\right|^{-2d}, \quad f^*(x) = \frac{\sigma_\xi^2}{2\pi} \left|1 - e^{ix}\right|^{-2d}.$$

In case of our simulations, where $\xi_k \sim N(0, 1)$ are i.i.d. standard normal innovations and $\varepsilon_k = \xi_k^2$, we have

$$\sigma^2 = 2, \quad \sigma_\xi^2 = 1,$$
$$\sigma_z^2(d = 0.01) = 2.0007, \quad \sigma_z^2(d = 0.10) = 2.0811,$$
$$\sigma_z^2(d = 0.20) = 2.4918, \quad \sigma_z^2(d = 0.30) = 5.4483,$$

and $\mu = 1$. So, in fact, the autocovariance and spectral density functions of FIGARCH model should dominate those of ARFIMA($0, d, 0$), while for the autocorrelation $\rho(k)$ we should have the same asymptotics in both cases:

$$\rho(k) \sim \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} k^{-1+2d}.$$

The empirical finding about the difference between ARFIMA($0, d, 0$) and FIGARCH ACFs is an indication that the simulation of the FIGARCH process is not a trivial task. Despite the fact that theoretically the condition (3.8) for the existence of the stationary FIGARCH process is satisfied, in practical simulations of times series models, whose setting is based on infinite past, various deviations from the theoretical model (e.g. the use of an initial condition $\tau_k = \mu, k \leq 0$) can have significant influence on the behavior of the resulting simulated paths, their ACFs, etc. One of possible reasons behind the unexpected difference between ARFIMA and FIGARCH covariance functions may be related to the length of the sample size. We repeated the simulation exercise using longer samples with $n = 40000$ and pre-sample $n_p = 10000$. Results show that ARFIMA ACF still dominates FIGARCH ACF, but differences become smaller, which means that the effect of increasing the sample size is bigger in the FIGARCH case and shifts the ACF upwards more than ARFIMA ACF (see Figure 3.5). However, it seems that the main problem is not the sample size.

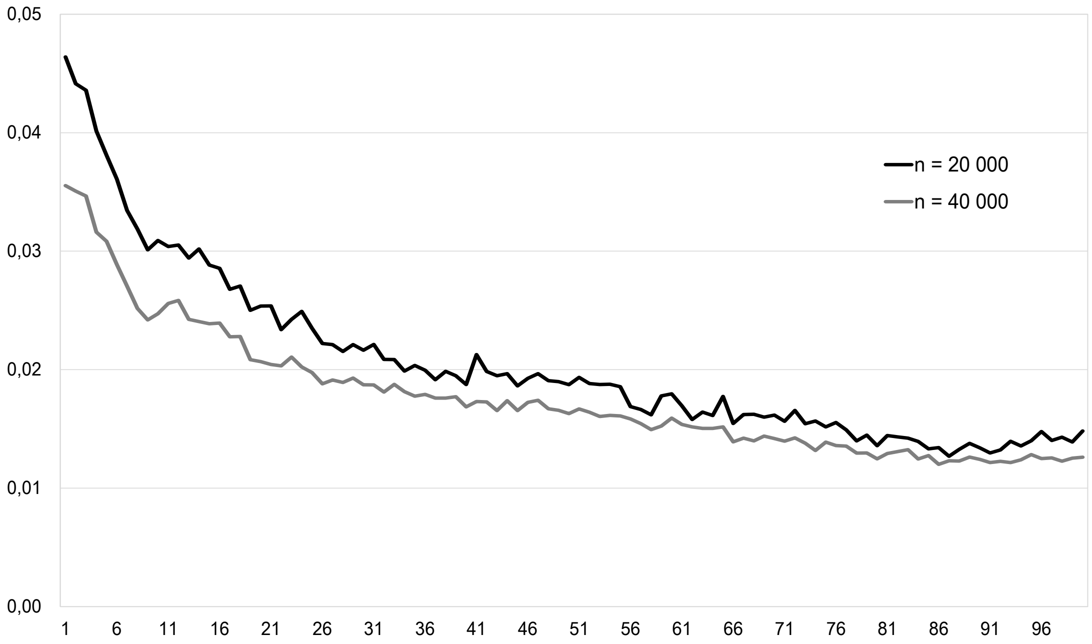

Figure 3.5: Differences between estimated ACF of ARFIMA$(0, d, 0)$ and FIGARCH processes for $d = 0.3$ with sample sizes $n = 40000$ and $n = 20000$ (lags $k = 1, ..., 100$; pre-sample $n_p = 10000$).

Further we check how this situation changes when the distribution of ξ_k is platykurtic, that is, when it does not produce large outliers which may have significant effect in practical simulations. We repeat the same simulation procedure with i.i.d. innovations ξ_k which are uniformly distributed over the interval $\left[-\sqrt{3}, \sqrt{3}\right]$. In this case, $\mathrm{E}\varepsilon_k = \mathrm{E}\xi_k^2 = 1$ and $\mathrm{E}\varepsilon_k^2 = \mathrm{E}\xi_k^4 = 1.8$. So now $\mathrm{E}\varepsilon_k^2$ is lower and more distant from the boundary of condition (3.8) for the existence of a stationary solution than in the case of standard normal innovations.

Figure 3.6 shows the corresponding ACFs of ARFIMA$(0, d, 0)$ and FIGARCH processes. In this case, differences between ARFIMA$(0, d, 0)$ and FIGARCH ACFs are obviously smaller and for $d < 0.3$ they almost disappear.

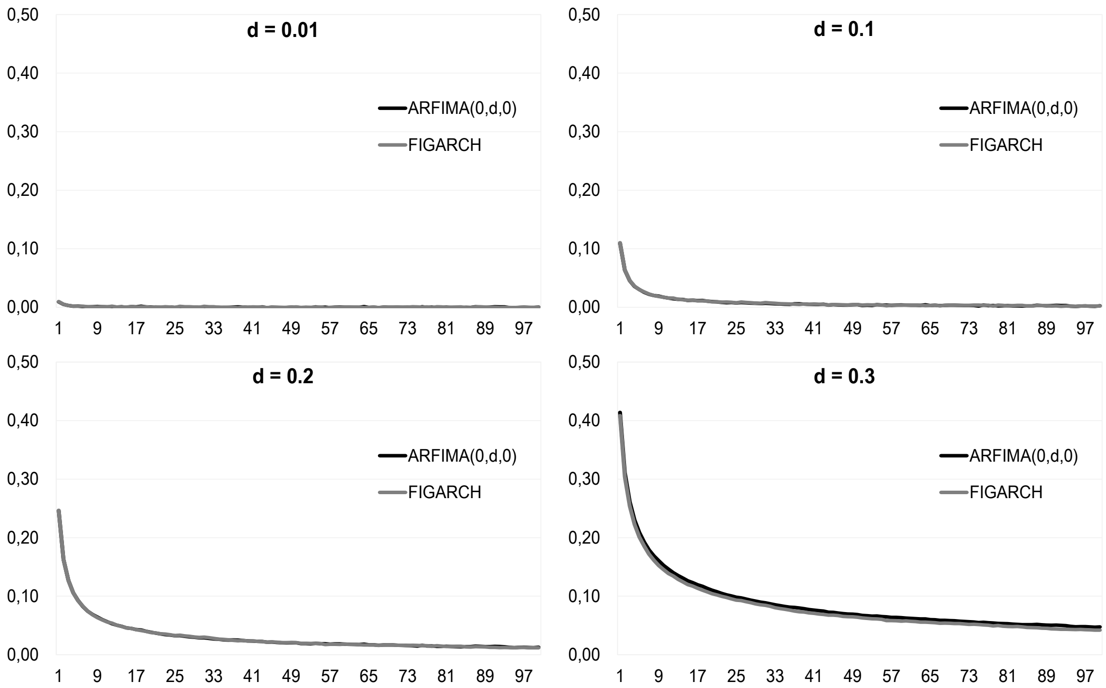

Figure 3.6: Estimated ACF of FIGARCH and ARFIMA$(0, d, 0)$ processes for different values of d (lags $k = 1, ..., 100$) with i.i.d. ζ_k uniformly distributed on $[-\sqrt{3}, \sqrt{3}]$ and $\varepsilon_k = \xi_k^2$.

Similar tendencies can be seen in terms of spectral densities. Figure 3.7 shows the estimated spectral densities of ARFIMA$(0, d, 0)$ and FIGARCH processes using: a) standard normal innovations and b) uniformly distributed innovations. In the first case, ARFIMA spectral density dominates that of FIGARCH near the origin. Our theoretical formulas show that the opposite situation should be in place. When innovations are distributed uniformly, the ARFIMA dominance decreases significantly.

These results indicate that in practical simulations, when one deals with finite-sample exercises, the difference between $\mathrm{E}\varepsilon_t^2$ and $\frac{\Gamma(1-2d)}{\Gamma(1-2d)-\Gamma^2(1-d)}$ should be "safe" enough to generate processes whose properties are in-line with theoretical findings.

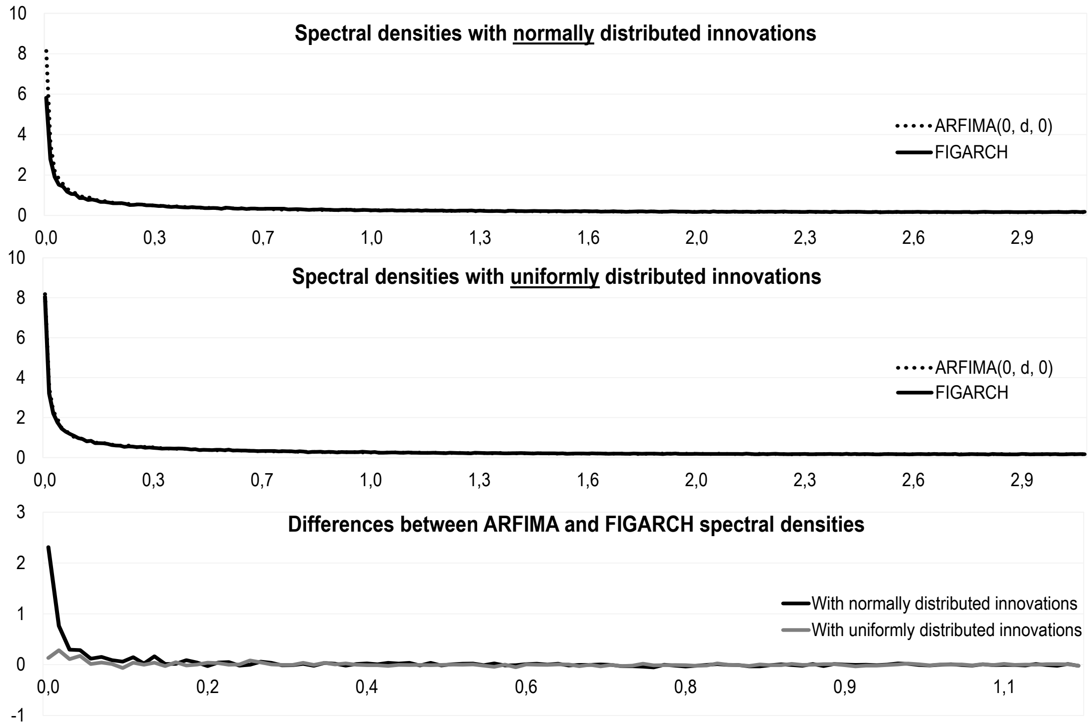

Figure 3.7: Estimated spectral densities of FIGARCH and ARFIMA$(0, d, 0)$ processes for $d = 0.3$ with i.i.d. ζ_k a) uniformly distributed on $[-\sqrt{3}, \sqrt{3}]$ and b) standard normal, and differences between spectral densities.

3.6.2 IAR(p, d, q) and ARFIMA(p, d, q) processes

In this subsection we explore through simulations the differences between the classical ARFIMA model

$$(1 - r_1 L - ... - r_p L^p)(1 - L)^d x_k = (1 + a_1 L + ... + a_q L^q)\xi_k, \quad k \in \mathbb{Z},$$

where $R(z) = 1 - r_1 z - ... - r_p z^p$ and $A(z) = 1 + a_1 z + ... + a_q z^q$, are polynomials of degrees $p, q \geq 0$, respectively, that have no common zeros, and the IAR(p, d, q) model

$$x_k = \sum_{j=1}^{\infty} b_j x_{k-j} + \xi_k + a_1 \xi_{k-1} + ... + a_q \xi_{k-q}. \tag{3.48}$$

Recall from Example 3.2 that coefficients b_j in IAR model (3.48) are generated by the operator

$$B(L) = \big(1-(1-L)^d\big)P(L) = \sum_{j=1}^{\infty} b_j L^j, \quad 0 < d < 1/2. \tag{3.49}$$

Here, $P(z) = \sum_{j=0}^{\infty} p_j z^j$ is a generating function with coefficients

$$p_j \geq 0, \quad p_0 > 0, \quad \sum_{j=0}^{\infty} p_j = 1 \quad \text{and} \quad \sum_{j=1}^{\infty} j p_j < \infty.$$

Then $b_j = \sum_{k=0}^{j-1} p_k b^0_{j-k}$, where b^0_j are coefficients of the expansion $1-(1-z)^d = \sum_{j=1}^{\infty} b^0_j z^j$ as in the previous subsection.

The IAR model is of particular interest, because, as mentioned in Example 3.2, the asymptotics of the IAR(p, d, q) covariance function is the same as for the ARFIMA($0, d, 0$) model, and the p_j or $P(L)$ in (3.49) essentially affects the short memory dynamics and do not distort the long-term behavior of the corresponding IAR process. This feature could be very interesting and useful from a practical point of view, as one can control the short-term behavior of a covariance function without an effect on the long-term asymptotics.

Figure 3.8 presents ACFs of ARFIMA($1, d, 0$)

$$(1 - r_1 L)(1-L)^d x_k = \xi_k,$$

and IAR($1, d, 0$)

$$x_k = \sum_{j=1}^{\infty} b_j x_{k-j} + \xi_k,$$

with $b_j = p_0 b^0_j + p_1 b^0_{j-1}$, and $1-(1-z)^d = \sum_{j=1}^{\infty} b^0_j z^j$. The simulation procedure is the same as in the previous section: the sample size is $n = 20000$ with the first $n_p = 10000$ values being the pre-sample, and theoretical ACFs are estimated using Monte Carlo averaging from $N = 500$ samples. In order to increase comparability between ARFIMA and IAR processes, we use the same innovations $\xi_k \sim N(0, 1)$ to get corresponding samples of both processes. The initial condition is $x_k = 1, k \leq 0$. The memory parameter is fixed at $d = 0.25$.

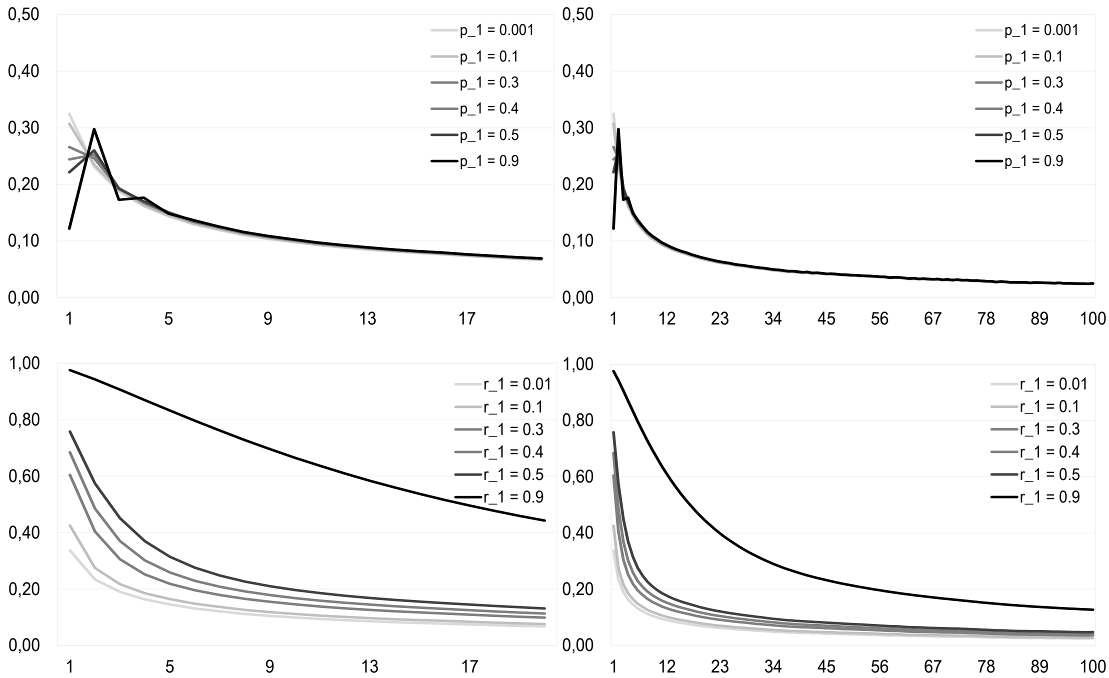

Figure 3.8: Estimated ACF of ARFIMA$(1, d, 0)$ (bottom panel) and IAR$(1, d, 0)$ (top panel) processes for $d = 0.25$ (lags $k = 1, ..., 20$ and $k = 1, ..., 100$).

From Figure 3.8 it is clear that these two models generate different covariance structures, except for very small $r_1 = p_0 = 0.01$. ARFIMA ACF is regularly decreasing in all cases, whereas for the IAR process we can achieve an ACF which is increasing at low lags, depending on the value of p_1. Estimated ACFs also indicate that changing the value of r_1 in

ARFIMA$(1, d, 0)$ has a different effect than p_1 in IAR$(1, d, 0)$. It seems that parameter p_1 influences the behavior of ACF only at a few lower lags, while r_1 creates a longer-lasting effect. The higher value of p_1 leads to lower ACF at lag $k = 1$ and increases the the value of ACF at lag $k = 2$.

However, in IAR$(1, d, 0)$, the maximum achievable peak of ACF at lag $k = 1$ seems to be quite low (about 0.35), while in practical applications it is desirable and natural to model processes with higher ACF at lower lags. This can be achieved using higher-order IAR(p, d, q) models. Before turning to models with $q > 0$, we present two examples, where $q = 0$ and $p = 2$ and $p = 3$.

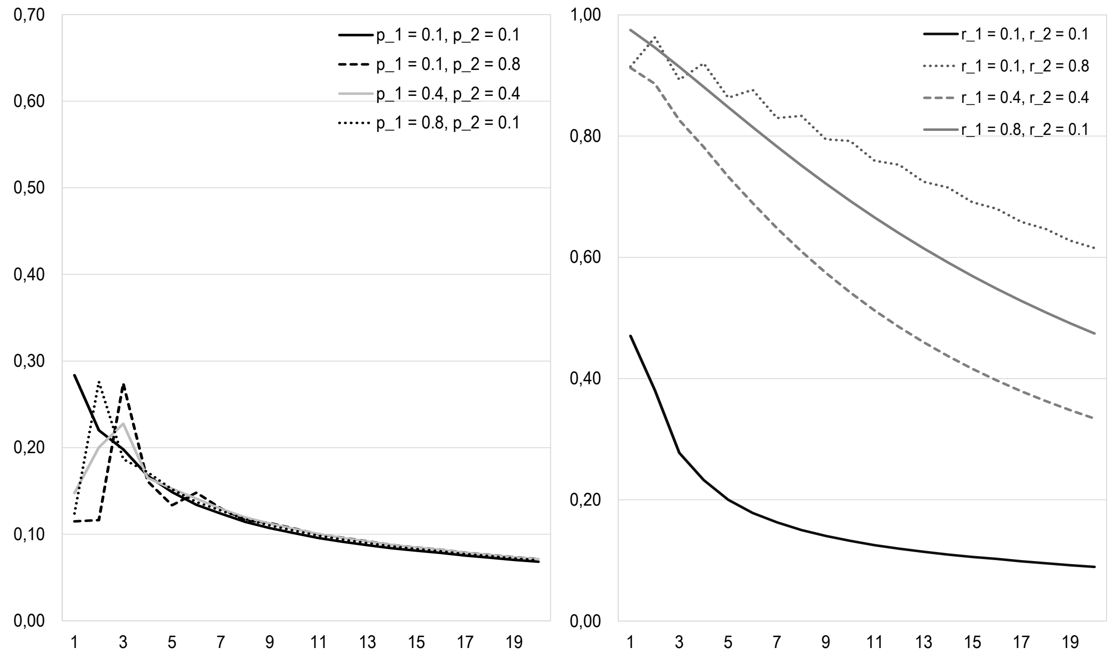

Figure 3.9: Estimated ACF of ARFIMA$(2, d, 0)$ (right panel) and IAR$(2, d, 0)$ (left panel) processes for $d = 0.25$ and different values of p_1, p_2, r_1, r_2 (lags $k = 1, ..., 20$).

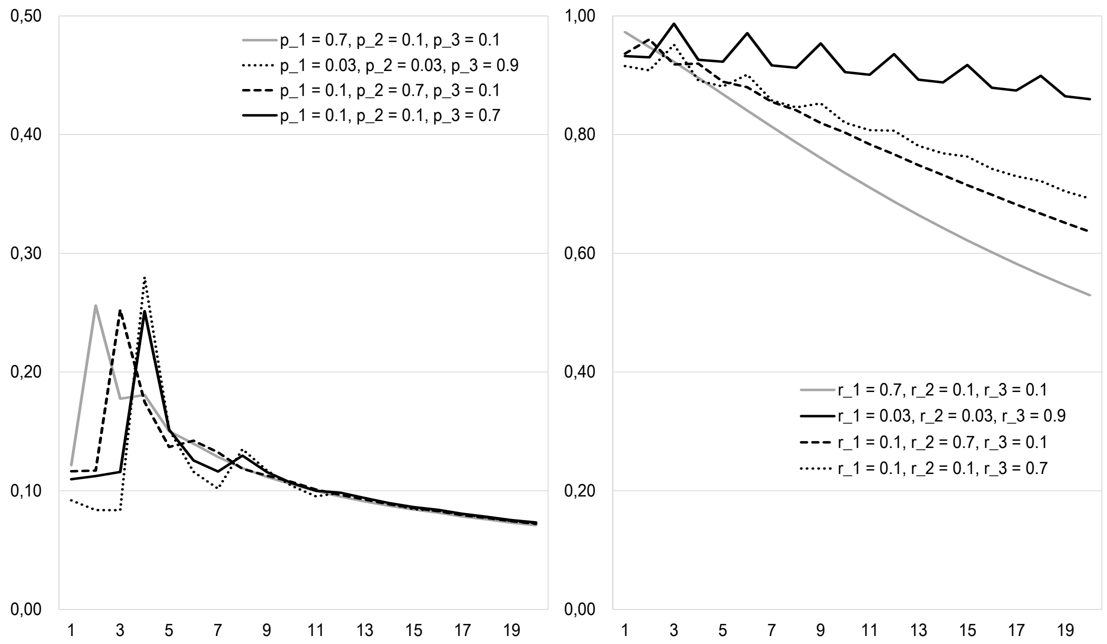

Figure 3.10: Estimated ACF of ARFIMA$(3, d, 0)$ (right panel) and IAR$(3, d, 0)$ (left panel) processes for $d = 0.25$ and different values of $p_1, p_2, p_3, r_1, r_2, r_3$ (lags $k = 1, ..., 20$).

From Figures 3.9 and 3.10 we can see how the choice of values for p_j in IAR$(p, d, 0)$ affects the behavior of ACF. High values of p_j generate the peak of ACF at lag $k = j + 1$. Choosing high values of p_0, we get the "traditional" regularly decreasing ACF. On the other hand, introducing nonzero values $r_j, j > 1$, in ARFIMA(p, d, q), changes the ARFIMA ACF drastically not only on low lags, but also in the long-run.

Figures 3.11–3.13 present ACFs of ARFIMA(p, d, q) and IAR(p, d, q) for $q = 1$ and $p = 1, 2, 3$. It seems that the main advantage of introducing the AR(1) component for ξ_k in the IAR setting is the increased ACF at lower lags without a major impact on ACF values in the long-run. At the same time we retain the possibility of controlling the peaks of ACF at low lags, changing the values of p_j.

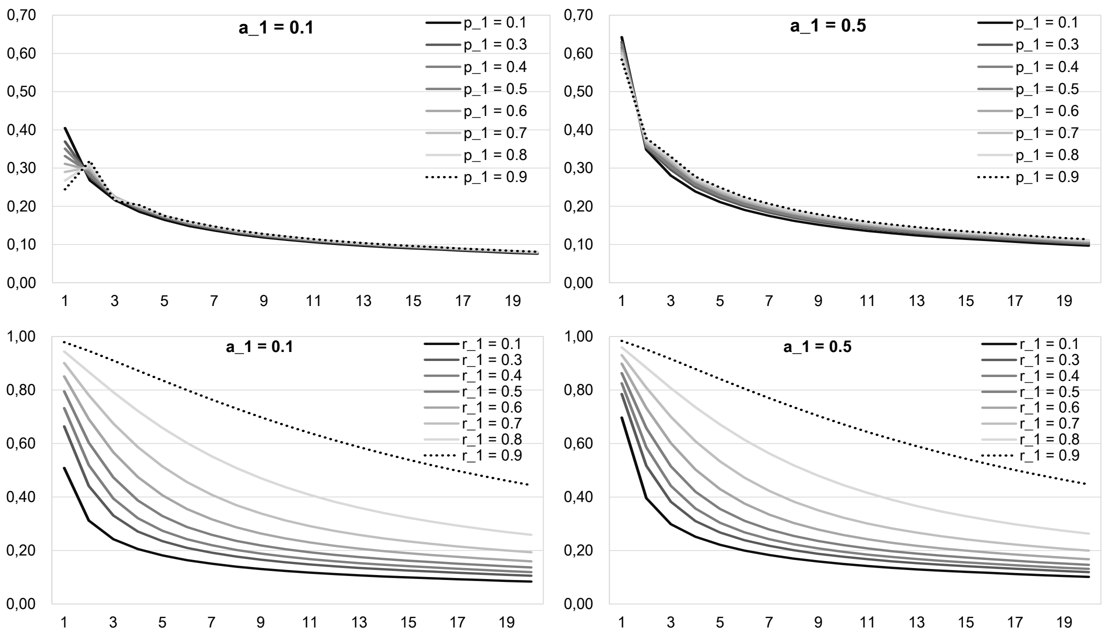

Figure 3.11: Estimated ACF of ARFIMA$(1, d, 1)$ (bottom panel) and IAR$(1, d, 1)$ (top panel) processes for $d = 0.25$ and different choices of p_1, r_1, a_1 (lags $k = 1, ..., 20$).

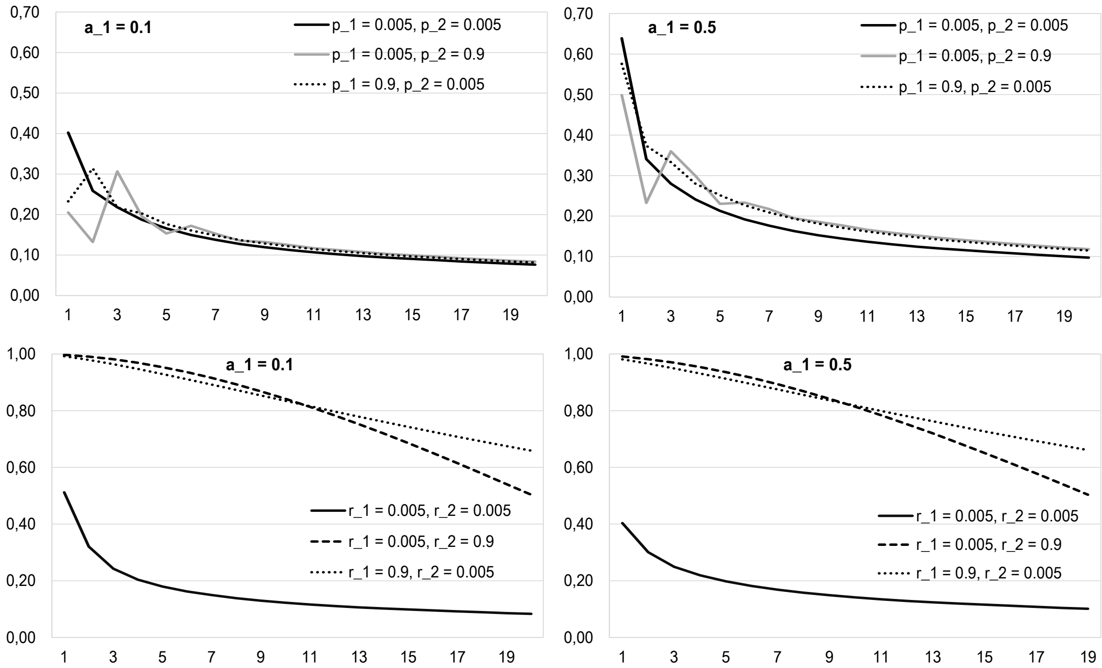

Figure 3.12: Estimated ACF of ARFIMA$(2, d, 1)$ (bottom panel) and IAR$(2, d, 1)$ (top panel) processes for $d = 0.25$ and different choices of p_1, p_2, r_1, r_2, a_1 (lags $k = 1, ..., 20$).

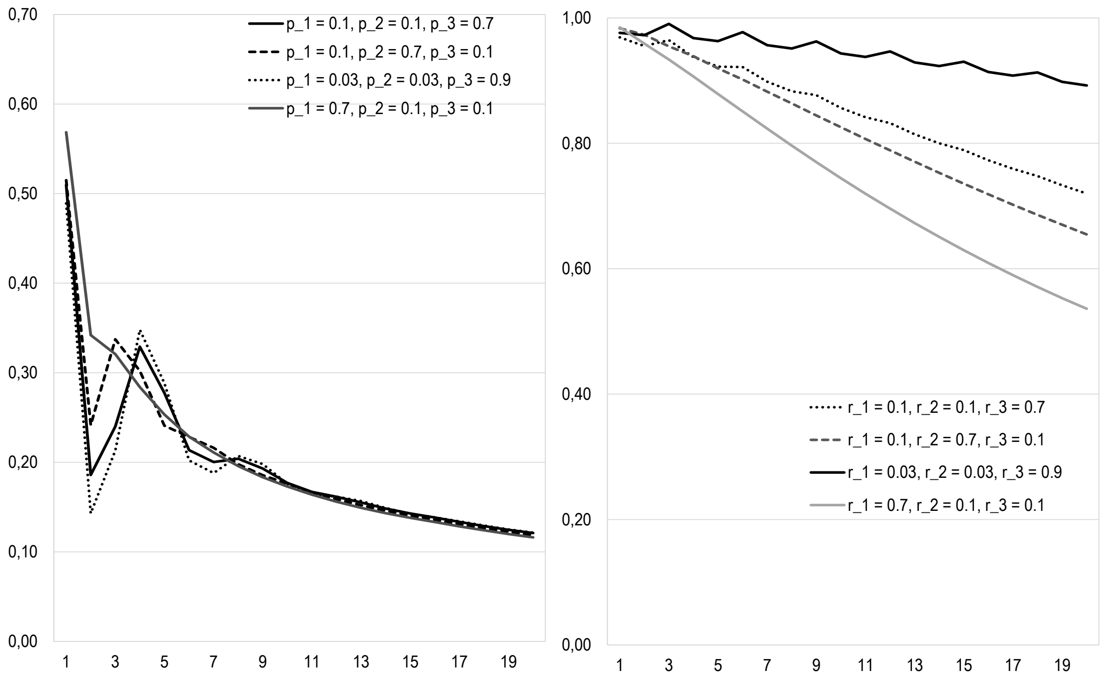

Figure 3.13: Estimated ACF of ARFIMA$(3, d, 1)$ (right panel) and IAR$(3, d, 1)$ (left panel) processes for $d = 0.25$ and different choices of $p_1, p_2, p_3, r_1, r_2, r_3$, $a_1 = 0.5$ (lags $k = 1, ..., 20$).

For practical purposes it would be very useful to find coefficients ψ_k of the infinite moving average representation of the IAR process

$$x_t = \psi(L)\xi_t = \sum_{j=0}^{\infty} \psi_j \xi_{t-j}.$$

Having the explicit form of ψ_j as well as covariance $\gamma(k)$ as well, one could use the Davies-Harte algorithm to generate the process with known (e.g. at the first k lags) covariance function. This algorithm was first discussed by Davies and Harte [19] (see also Giraitis *et al.* [36]).

We try to find coefficients ψ_j for IAR$(1, d, 0)$. Since

$$(1 - B(L))x_t = (1 - (1 - (1 - L)^d)P(L))x_t = \xi_t,$$

writing

$$x_t = (1 - B(L))^{-1}\xi_t,$$

we get

$$(1 - B(L))^{-1} = \sum_{k=0}^{\infty} B^k(L) = \sum_{k=0}^{\infty}(1-(1-L)^d)^k P^k(L) =$$
$$= \sum_{k=0}^{\infty}\left[\sum_{s=0}^{k}\binom{k}{s}(-1)^s\left(\sum_{n=0}^{\infty}\frac{ds(ds-1)\cdot\ldots\cdot(ds-n+1)}{n!}(-1)^n L^n\right)\right]\cdot$$
$$\cdot\left[\sum_{i=0}^{k}\binom{k}{i}p_0^{k-i}p_1^i L^i\right] = \sum_{k=0}^{\infty}\left[\sum_{s=0}^{k}\binom{k}{s}(-1)^s\left(1 - dsL + \frac{ds(ds-1)}{2}L^2\right.\right.$$
$$\left.\left.-\ \frac{ds(ds-1)(ds-2)}{3!}L^3 + \ldots\right)\right]\cdot\left[p_0^k + \frac{k!}{(k-1)!}p_0^{k-1}p_1 L + \ldots\right.$$
$$\left.\ldots + p_1^k L^k\right] = \sum_{k=0}^{\infty}\left[\sum_{s=0}^{k}\binom{k}{s}(-1)^s + L\sum_{s=0}^{k}\binom{k}{s}(-1)^{s+1}ds + \right.$$
$$\left. + L^2\sum_{s=0}^{k}\binom{k}{s}(-1)^{s+2}\frac{ds(ds-1)}{2} + \ldots\right]\cdot$$
$$\cdot\left[p_0^k + \frac{k!p_0^{k-1}}{(k-1)!}p_1 L + \ldots + p_1^k L^k\right].$$

Collecting members at different powers of L for fixed $k \in \{0, 1, 2, ...\}$ we have

$$p_0^k\sum_{s=0}^{k}\binom{k}{s}(-1)^s +$$
$$L\left(p_0^k\sum_{s=0}^{k}\binom{k}{s}(-1)^{s+1}ds + \binom{k}{1}p_0^{k-1}p_1\sum_{s=0}^{k}\binom{k}{s}(-1)^s\right) +$$
$$L^2\left(p_0^k\sum_{s=0}^{k}\binom{k}{s}(-1)^{s+2}\frac{ds(ds-1)}{2} + \binom{k}{1}p_0^{k-1}p_1\sum_{s=0}^{k}\binom{k}{s}(-1)^{s+1}ds + \right.$$
$$\left. + \binom{k}{2}p_0^{k-2}p_1^2\sum_{s=0}^{k}\binom{k}{s}(-1)^s\right) +$$

$$
\begin{aligned}
& L^3 \left(p_0^k \sum_{s=0}^{k} \binom{k}{s} (-1)^{s+3} \frac{ds(ds-1)(ds-2)}{3!} + \right. \\
& + \binom{k}{3} p_0^{k-3} p_1^3 \sum_{s=0}^{k} \binom{k}{s} (-1)^s + \binom{k}{2} p_0^{k-2} p_1^2 \sum_{s=0}^{k} \binom{k}{s} ds(-1)^{s+1} + \\
& \left. + \binom{k}{1} p_0^{k-1} p_1 \sum_{s=0}^{k} \binom{k}{s} (-1)^{s+2} \frac{ds(ds-1)}{2} \right) + \ldots = \\
= & \sum_{m=0}^{\infty} L^m \left(\sum_{s=0}^{m} \binom{k}{s} p_0^{k-s} p_1^s \left[\sum_{j=0}^{k} \binom{k}{j} \binom{dj}{m-s} (-1)^{j+m-s} \right] \right) = \\
= & \sum_{m=0}^{\infty} L^m \left(\sum_{s=0}^{m} \binom{k}{s} \left[\sum_{z=0}^{k-s} \binom{k-s}{z} (-1)^z p_1^{z+s} \right] \cdot \right. \\
& \left. \cdot \left[\sum_{j=0}^{k} \binom{k}{j} \binom{dj}{m-s} (-1)^{j+m-s} \right] \right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
(1-B(L))^{-1} &= \sum_{m=0}^{\infty} \psi_m L^m \\
&= \sum_{m=0}^{\infty} L^m \sum_{k=0}^{\infty} \left(\sum_{s=0}^{m} \binom{k}{s} \left[\sum_{z=0}^{k-s} \binom{k-s}{z} (-1)^z p_1^{z+s} \right] \right. \\
&\quad \left. \cdot \left[\sum_{j=0}^{k} \binom{k}{j} \binom{dj}{m-s} (-1)^{j+m-s} \right] \right).
\end{aligned}
$$

Since

$$
\sum_{j=0}^{k} \binom{k}{j} (-1)^j P(j) = 0,
$$

for polynomial $P(x)$ of degree $s < k$, coefficient ψ_m can be truncated to

$$
\psi_m = \sum_{k=0}^{m} \left(\sum_{s=0}^{\min(k,m-k)} \binom{k}{s} p_0^{k-s} p_1^s \left[(-1)^{m-s} \sum_{j=0}^{k} \binom{k}{j} \binom{dj}{m-s} (-1)^j \right] \right). \tag{3.50}
$$

Explicit expressions of the first four coefficients ψ_m are:

$$\begin{aligned}\psi_0 &= 1, \qquad \psi_1 = p_0 d, \qquad \psi_2 = p_1 d + p_0 d\left(\frac{1}{2}(1-d) + p_0 d\right),\\ \psi_3 &= dp_0\left(d^2p_0^2 - d^2p_0 + \frac{d^2}{6} + -dp_0 + 2d - \frac{1}{6}\right) + \frac{d(1-d)}{2}.\end{aligned}$$

Figure 3.14 contains coefficients $\psi_k, k = 1, ..., 20$, calculated using 3.50 for $d = 0.3$ and different values of parameter p_1. Unfortunately, formula (3.50) is inconvenient for practical purposes since it is very time consuming for larger m (e.g., $m > 50$). We believe that there is room for simplification of (3.50), yet it is left for our future work.

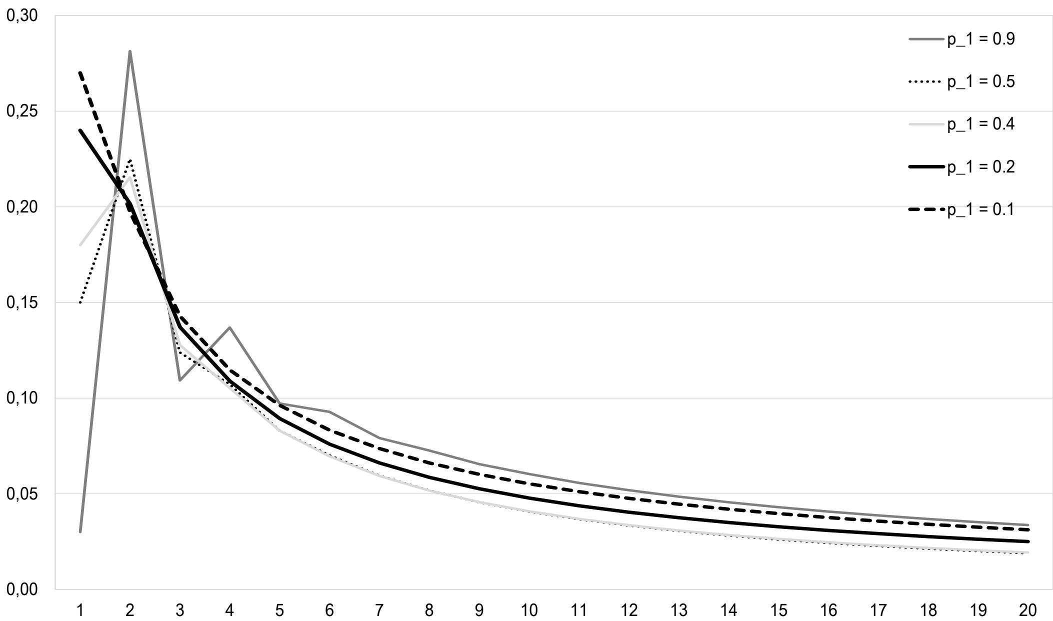

Figure 3.14: Coefficients $\psi_k, k = 1, ..., 20$ from (3.50) with $d = 0.3$.

Here we end the simulation study and present the main conclusions of the chapter.

3.7 Conclusion

Ding and Granger [20] proposed the Long Memory ARCH model to capture the hyperbolic decay of sample autocorrelations of speculative squared returns. The LM ARCH model is closely related to the FIGARCH model which was independently introduced by Baillie *et al.* [3]. However, the existence of a covariance stationary solution of these models was not established and, thus, the possibility of long memory in the ARCH setting was doubtful. In this dissertation, we solved this controversy by showing that FIGARCH and IARCH(∞) equations with zero intercept may have a nontrivial covariance stationary solution with long memory. We also obtained necessary and sufficient conditions for the existence of stationary integrated AR(∞) processes with finite variance and proved that such processes always have long memory. We provided a complete answer to the long standing conjecture of Ding and Granger [20] about the existence of the Long Memory ARCH model.

Chapter 4

Quasi-MLE for the quadratic ARCH model with long memory

We discuss the parametric quasi-maximum likelihood estimation for the quadratic ARCH (QARCH) process with long memory, introduced by Doukhan, Grublytė, and Surgailis [22] and Grublytė and Škarnulis [40] (see also Chapter 5 of this dissertation), with conditional variance involving the square of inhomogeneous linear combination of an observable sequence with square summable weights. The above model extends the QARCH model of Sentana [66] and the Linear ARCH model of Robinson [62] to the case of strictly positive conditional variance. We prove consistency and asymptotic normality of the corresponding QML estimates, including the estimate of the long memory parameter $0 < d < 1/2$. A simulation study of empirical MSE is included.

4.1 Introduction

Recently, Doukhan *et al.* [22] and Grublytė and Škarnulis [40] discussed a class of quadratic ARCH models of the form

$$r_t \;=\; \zeta_t\sigma_t, \qquad \sigma_t^2 \;=\; \omega^2 + \Big(a + \sum_{j=1}^{\infty} b_j r_{t-j}\Big)^2 + \gamma\sigma_{t-1}^2, \tag{4.1}$$

where $\{\zeta_t, t \in \mathbb{Z}\}$ is a standardized i.i.d. sequence, $\mathrm{E}\zeta_t = 0, \mathrm{E}\zeta_t^2 = 1$, and $\gamma \in [0,1), \omega, a, b_j, j \geq 1$, are real parameters satisfying certain conditions, see Proposition 4.1 below. Grublytė and Škarnulis [40] called (4.1) the Generalized Quadratic ARCH (GQARCH) model. It is considered in more detail way in Chapter 5 of this dissertation. By iterating the second equation in (4.1), the squared volatility in (4.1) can be written as a quadratic form

$$\sigma_t^2 \;=\; \sum_{\ell=0}^{\infty} \gamma^\ell \Big\{\omega^2 + \big(a + \sum_{j=1}^{\infty} b_j r_{t-\ell-j}\big)^2\Big\}$$

in lagged variables $r_{t-1}, r_{t-2}, ...$, and hence it represents a particular case of Sentana's [66] Quadratic ARCH model with $p = \infty$. The model (4.1) includes the classical Asymmetric GARCH(1,1) process of Engle [25] and the Linear ARCH (LARCH) model of Robinson [62]:

$$r_t \;=\; \zeta_t\sigma_t, \qquad \sigma_t \;=\; a + \sum_{j=1}^{\infty} b_j r_{t-j}. \tag{4.2}$$

The main interest in (4.1) and (4.2) seems to be the possibility of having slowly decaying moving-average coefficients b_j with $\sum_{j=1}^{\infty}|b_j| = \infty, \sum_{j=1}^{\infty} b_j^2 < \infty$, for modeling long memory in volatility, in which case, r_t and ζ_t must have zero mean so that the series $\sum_{j=1}^{\infty} b_j r_{t-j}$ converges.

Giraitis *et al.* [32] proved that the squared stationary solution $\{r_t^2, t \in \mathbb{Z}\}$ of the LARCH model in (4.2) with b_j decaying as $j^{d-1}, 0 < d < 1/2$, may have long memory autocorrelations. In terms of the GQARCH model in (4.1), similar results were established by Doukhan *et al.* [22] and Grublytė and Škarnulis [40]. Namely, assume that parameters $\gamma, \omega, a, b_j, j \geq 1$, in (4.1) satisfy

$$b_j \ \sim \ c\, j^{d-1} \qquad (\exists\, 0 < d < 1/2,\ c > 0),$$

$\gamma \in [0, 1),\ a \neq 0$ and

$$6B_2 + 4|\mu_3| \sum_{j=1}^{\infty} |b_j|^3 + \mu_4 \sum_{j=1}^{\infty} b_j^4 \ < \ (1-\gamma)^2, \tag{4.3}$$

where $\mu_p := \mathrm{E}\zeta_0^p,\ p = 1, 2, ...,\ B_2 := \sum_{j=1}^{\infty} b_j^2$. Then (Grublytė and Škarnulis [40], Theorems 2.5 and 3.1) there exists a stationary solution of (4.1) with $\mathrm{E}r_t^4 < \infty$, such that

$$\mathrm{cov}(r_0^2, r_t^2) \ \sim \ \kappa_1^2 t^{2d-1}, \qquad t \to \infty,$$

and

$$n^{-d-1/2} \sum_{t=1}^{[ns]} (r_t^2 - \mathrm{E}r_t^2) \ \to_{D[0,1]} \ \kappa_2 B_{d+(1/2)}(s), \qquad n \to \infty,$$

where $B_{d+(1/2)}$ is a fractional Brownian motion with the Hurst parameter $H = d + (1/2) \in (1/2, 1)$ (see also Definition 2.11) and $\kappa_i > 0, i = 1, 2$, are some constants; $\to_{D[0,1]}$ stands for weak convergence in the Skorohod space $D[0, 1]$.

As noted by Doukhan *et al.* [22] and Grublytė and Škarnulis [40], the GQARCH model of (4.1) and the LARCH model of (4.2) have similar long memory and leverage properties and can both be used for modeling financial data with the above properties. The main disadvantage of the latter model in comparison to the former one seems to be the fact that volatility σ_t in (4.2) may assume negative values and is not separated from below by positive constant $c > 0$ as in the case of (4.1). The standard quasi-maximum likelihood (QML) approach to the estimation of LARCH parameters is inconsistent and other estimation methods were developed by Beran and Schützner [5], Francq and Zakoian [29], Levine *et al.* [52], Truquet [68]. The results of Doukhan *et al.* [22] and Grublytė and Škarnulis [40] are limited to properties of the stationary solution of (4.1) and do not include estimation or other statistical inferences for this model.

In this chapter of the dissertation, we discuss the QML estimation for the 5-parametric GQARCH model

$$\sigma_t^2(\theta) \;=\; \sum_{\ell=0}^{\infty} \gamma^{\ell} \left\{ \omega^2 + \big(a + c\sum_{j=1}^{\infty} j^{d-1} r_{t-\ell-j}\big)^2 \right\}, \tag{4.4}$$

depending on parameter $\theta = (\gamma, \omega, a, d, c), 0 < \gamma < 1, \omega > 0, a \neq 0, c \neq 0$ and $d \in (0, 1/2)$. The parametric form $b_j = c\, j^{d-1}$ of moving-average coefficients in (4.4) is the same as that by Beran and Schützner [5] for the LARCH model. Similar to Beran and Schützner [5], we discuss the QML estimator

$$\widehat{\theta}_n := \arg\min_{\theta\in\Theta} L_n(\theta), \quad L_n(\theta) := \frac{1}{n}\sum_{t=1}^{n}\left(\frac{r_t^2}{\sigma_t^2(\theta)} + \log\sigma_t^2(\theta)\right),$$

involving exact conditional variance (4.4) depending on infinite past $r_s, -\infty < s < t$, and its more realistic version $\widetilde{\theta}_n := \arg\min_{\theta\in\Theta} \widetilde{L}_n(\theta)$, obtained by replacing the $\sigma_t^2(\theta)$'s in (4.4) by $\widetilde{\sigma}_t^2(\theta)$ depending only $r_s, 1 \leq s < t$ (see Section 4.3 for the definition). It should be noted that the QML function proposed by Beran and Schützner [5] is modified to avoid the degeneracy of σ_t^{-1} in (4.2), by introducing an additional tuning parameter $\epsilon > 0$ which affects the performance of the estimator and whose choice is a nontrivial task. In terms of the GQARCH model (4.4) with $\omega > 0$, the aforementioned degeneracy problem does not occur and we deal with unmodified QMLE in contrast to Beran and Schützner [5]. We also note that our proofs use techniques different from those of Beran and Schützner [5]. Particularly, the method of orthogonal Volterra expansions of the LARCH model used by Beran and Schützner [5] is not applicable for model (4.4); see Doukhan *et. al.* ([22], Example 1).

Section 4.2 presents some results of Grublytė and Škarnulis [40] about the existence and properties of the stationary solution of GQARCH equations in (4.1). More details about the GQARCH process are provided in Chapter 5 of this dissertation. In Section 4.3, we define several QML estimators of parameter θ in (4.4). Section 4.4 presents the main results of this chapter related to consistency and asymptotic normality of the QML estimators. Finite sample performance of these estimators is investigated in the simulation study in Section 4.5. Conclusions are summarized in Section 4.7 and all proofs are relegated to Section 4.6.

4.2 Stationary solution

We recall a few facts from Chapter 5 of this dissertation about the stationary solution of (4.1) (see also Grublytė and Škarnulis [40]). First, we give its definition. Let $\mathcal{F}_t = \sigma(\zeta_s, s \le t), t \in \mathbb{Z}$, be the sigma-field generated by $\zeta_s, s \le t$.

Definition 4.1. *By stationary solution of (4.1) we mean a stationary and ergodic martingale difference sequence* $\{r_t, \mathcal{F}_t, t \in \mathbb{Z}\}$ *with* $\mathrm{E}r_t^2 < \infty, \mathrm{E}[r_t^2|\mathcal{F}_{t-1}] = \sigma_t^2$*, such that for any* $t \in \mathbb{Z}$ *the series* $X_t := \sum_{s<t} b_{t-s} r_s$ *converges in* L^2*, the series* $\sigma_t^2 = \sum_{\ell=0}^{\infty} \gamma^\ell (\omega^2 + (a + X_{t-\ell})^2)$ *converges in* L^1 *and (4.1) holds.*

Proposition 4.1. (Grublytė and Škarnulis [40]) *Let* $\gamma \in [0, 1)$ *and* $\{\zeta_t, t \in \mathbb{Z}\}$ *be an i.i.d. sequence with zero mean, unit variance and finite moment* $\mu_p := \mathrm{E}\zeta_0^p < \infty$*, where* $p \geq 2$ *is an even integer. Assume that*

$$\sum_{j=2}^{p} \binom{p}{j} |\mu_j| \sum_{k=1}^{\infty} |b_k|^j \;<\; (1-\gamma)^{p/2}. \tag{4.5}$$

Then there exists a unique stationary solution $\{r_t, t \in \mathbb{Z}\}$ *of (4.1) such that the series* $X_t = \sum_{j=1}^{\infty} b_j r_{t-j}$ *converges in* L^p *and* $\mathrm{E}r_t^p \leq C(1 + \mathrm{E}X_t^p) < \infty$*. Moreover, for* $p = 2$*, condition (4.5), or*

$$B_2 = \sum_{j=1}^{\infty} b_j^2 < 1 - \gamma, \tag{4.6}$$

is necessary and sufficient for the existence of a stationary L^2*-solution of (4.1) with*

$$\mathrm{E}r_t^2 = \frac{\omega^2 + a^2}{1 - \gamma - B_2}.$$

Remark 4.1. Condition (4.5) coincides with the corresponding condition for the LARCH model obtained by Giraitis *et al.* ([33], Proposition 3). For $p = 4$ (4.5) agrees with (4.3).

Remark 4.2. Sufficient conditions for the existence of a stationary solution of (4.1) with finite moment $\mathrm{E}|r_t|^p < \infty$ and arbitrary $p > 0$ are obtained in Chapter 5 of this dissertation, Theorem 5.1 (see also Grublytė and

Škarnulis [40], Theorem 2.4). There we extend the corresponding result of Doukhan *et al.* ([22], Theorem 1) from $\gamma = 0$ to $\gamma > 0$. Contrary to (4.5), the above-mentioned conditions involve absolute constant K_p from the Burkholder-Rosenthal inequality, which is not known explicitly, and, therefore, these conditions are not very useful (see Remark 5.2 in Chapter 5 of this dissertation; also Grublytė and Škarnulis ([40]).

4.3 QML Estimators

The following assumptions on the parametric GQARCH model in (4.4) are imposed.

Assumption (A) $\{\zeta_t, t \in \mathbb{Z}\}$ is a standardized i.i.d. sequence with $\mathrm{E}\zeta_t = 0, \mathrm{E}\zeta_t^2 = 1$.

Assumption (B) $\Theta \subset \mathbb{R}^5$ is a compact set of parameters $\theta = (\gamma, \omega, a, d, c)$ defined by

(i) $\gamma \in [\gamma_1, \gamma_2]$ with $0 < \gamma_1 < \gamma_2 < 1$;

(ii) $\omega \in [\omega_1, \omega_2]$ with $0 < \omega_1 < \omega_2 < \infty$;

(iii) $a \in [a_1, a_2]$ with $-\infty < a_1 < a_2 < \infty$;

(iv) $d \in [d_1, d_2]$ with $0 < d_1 < d_2 < 1/2$;

(v) $c \in [c_1, c_2]$ with $0 < c_i = c_i(d, \gamma) < \infty,\ c_1 < c_2$ such that

$$B_2 = c^2 \sum_{j=1}^{\infty} j^{2(d-1)} < 1 - \gamma$$

for any $c \in [c_1, c_2], \gamma \in [\gamma_1, \gamma_2], d \in [d_1, d_2]$.

We assume that the observations $\{r_t, 1 \le t \le n\}$ follow the model in (4.1) with the true parameter $\theta_0 = (\gamma_0, \omega_0, a_0, d_0, c_0)$, belonging to the interior Θ_0 of Θ in Assumption (B). The restriction on parameter c in (v) is due to condition (4.6). The QML estimator of $\theta \in \Theta$ is defined as

$$\widehat{\theta}_n := \arg\min_{\theta \in \Theta} L_n(\theta), \tag{4.7}$$

where

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{r_t^2}{\sigma_t^2(\theta)} + \log \sigma_t^2(\theta) \right), \tag{4.8}$$

and $\sigma_t^2(\theta)$ is defined in (4.4), that is,

$$\begin{aligned} \sigma_t^2(\theta) &= \sum_{\ell=0}^{\infty} \gamma^{\ell} \left\{ \omega^2 + \big(a + cY_{t-\ell}(d)\big)^2 \right\}, \qquad \text{where} \\ Y_t(d) &:= \sum_{j=1}^{\infty} j^{d-1} r_{t-j}. \end{aligned} \tag{4.9}$$

Note that the definitions in (4.7)–(4.9) depend on (unobserved) $r_s, s \le 0$, and, therefore, the estimator in (4.7) is usually referred to as the QML estimator given infinite past (see Beran and Schützner [5]). A more realistic version of (4.7) is defined as

$$\widetilde{\theta}_n := \arg\min_{\theta \in \Theta} \widetilde{L}_n(\theta), \tag{4.10}$$

where

$$\begin{aligned} \widetilde{L}_n(\theta) &:= \frac{1}{n} \sum_{t=1}^{n} \left(\frac{r_t^2}{\widetilde{\sigma}_t^2(\theta)} + \log \widetilde{\sigma}_t^2(\theta) \right), \quad \text{where} \\ \widetilde{\sigma}_t^2(\theta) &:= \sum_{\ell=0}^{t-1} \gamma^{\ell} \left\{ \omega^2 + \big(a + c\widetilde{Y}_{t-\ell}(d)\big)^2 \right\}, \quad \widetilde{Y}_t(d) := \sum_{j=1}^{t-1} j^{d-1} r_{t-j}. \end{aligned} \tag{4.11}$$

Note that all quantities in (4.11) depend only on $r_s, 1 \le s \le n$; hence (4.10) is called the QML estimator given finite past. The QML functions in (4.8) and (4.11) can be written as

$$L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta) \quad \text{and} \quad \widetilde{L}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{l}_t(\theta),$$

respectively, where

$$l_t(\theta) := \frac{r_t^2}{\sigma_t^2(\theta)} + \log \sigma_t^2(\theta), \qquad \widetilde{l}_t(\theta) := \frac{r_t^2}{\widetilde{\sigma}_t^2(\theta)} + \log \widetilde{\sigma}_t^2(\theta). \tag{4.12}$$

Finally, following Beran and Schützner [5] we define a truncated version of (4.10), involving the last $O(n^\beta)$ quasi-likelihoods $\widetilde{l}_t(\theta), n - [n^\beta] < t \le n$, as follows:

$$\widetilde{\theta}_n^{(\beta)} := \arg\min_{\theta\in\Theta} \widetilde{L}_n^{(\beta)}(\theta), \qquad \widetilde{L}_n^{(\beta)}(\theta) := \frac{1}{[n^\beta]} \sum_{t=n-[n^\beta]+1}^{n} \widetilde{l}_t(\theta), \tag{4.13}$$

where $0 < \beta < 1$ is a "bandwidth parameter". Note that for any $t \in \mathbb{Z}$ and $\theta_0 = (\gamma_0, \omega_0, a_0, d_0, c_0) \in \Theta$, the random functions $Y_t(d)$ and $\widetilde{Y}_t(d)$ in (4.9) and (4.11) are infinitely differentiable w.r.t. $d \in (0, 1/2)$ a.s. Hence, using the explicit form of $\sigma_t^2(\theta)$ and $\widetilde{\sigma}_t^2(\theta)$, it follows that $\sigma_t^2(\theta)$, $\widetilde{\sigma}_t^2(\theta)$, $l_t(\theta)$, $\widetilde{l}_t(\theta)$, $L_n(\theta)$, $\widetilde{L}_n(\theta)$, $\widetilde{L}_n^{(\beta)}(\theta)$ and so on, are all infinitely differentiable w.r.t. $\theta \in \Theta_0$ a.s. We use the notation

$$L(\theta) := \mathrm{E} L_n(\theta) = \mathrm{E} l_t(\theta), \tag{4.14}$$

and

$$A(\theta) := \mathrm{E}\left[\nabla^T l_t(\theta) \nabla l_t(\theta)\right], \quad B(\theta) := \mathrm{E}\left[\nabla^T \nabla l_t(\theta)\right], \tag{4.15}$$

where $\nabla = (\partial/\partial\theta_1, \cdots, \partial/\partial\theta_5)$ and the superscript "T" stands for transposed vector. Particularly, $A(\theta)$ and $B(\theta)$ are 5×5-matrices. By Lemma 4.1, the expectations in (4.15) are well-defined for any $\theta \in \Theta$ under condition $\mathrm{E}r_0^4 < \infty$. We have

$$B(\theta) \;=\; \mathrm{E}[\sigma_t^{-4}(\theta)\nabla^T\sigma_t^2(\theta)\nabla\sigma_t^2(\theta)] \quad \text{and} \quad A(\theta) = \kappa_4 B(\theta), \tag{4.16}$$

where $\kappa_4 := \mathrm{E}(\zeta_0^2 - 1)^2 > 0$.

4.4 Main results

Everywhere in this section $\{r_t, t \in \mathbb{Z}\}$ is a stationary solution of model (4.4) as defined in Definition 4.1 and satisfying Assumptions (A) and (B) of the previous section. As usual, all expectations are taken with respect to the true value $\theta_0 = (\gamma_0, \omega_0, a_0, d_0, c_0) \in \Theta_0$, where Θ_0 is the interior of the parameter set $\Theta \subset \mathbb{R}^5$.

Theorem 4.1. (i) *Let* $\mathrm{E}|r_t|^3 < \infty$. *Then* $\widehat{\theta}_n$ *in (4.7) is a strongly consistent estimator of* θ_0*, that is,*

$$\widehat{\theta}_n \overset{a.s.}{\to} \theta_0.$$

(ii) *Let* $\mathrm{E}|r_t|^5 < \infty$. *Then* $\widehat{\theta}_n$ *in (4.7) is asymptotically normal:*

$$n^{1/2}\big(\widehat{\theta}_n - \theta_0\big) \overset{d}{\to} N(0, \Sigma(\theta_0)), \tag{4.17}$$

where $\Sigma(\theta_0) := B^{-1}(\theta_0)A(\theta_0)B^{-1}(\theta_0) = \kappa_4 B^{-1}(\theta_0)$ *and matrices* $A(\theta), B(\theta)$ *are defined in (4.16).*

The following theorem gives asymptotic properties of the "finite past" estimators $\widetilde{\theta}_n$ and $\widetilde{\theta}_n^{(\beta)}$ defined in (4.10) and (4.13), respectively.

Theorem 4.2. (i) *Let* $\mathrm{E}|r_t|^3 < \infty$ *and* $0 < \beta < 1$. *Then*

$$\mathrm{E}|\widetilde{\theta}_n - \theta_0| \ \to\ 0 \qquad \textit{and} \qquad \mathrm{E}|\widetilde{\theta}_n^{(\beta)} - \theta_0| \ \to\ 0.$$

(ii) *Let* $\mathrm{E}|r_t|^5 < \infty$ *and* $0 < \beta < 1 - 2d_0$. *Then*

$$n^{\beta/2}(\widetilde{\theta}_n^{(\beta)} - \theta_0) \ \stackrel{d}{\to}\ N(0, \Sigma(\theta_0)), \tag{4.18}$$

where $\Sigma(\theta_0)$ *is the same as in Theorem 4.1.*

The asymptotic results in Theorems 4.1 and 4.2 are similar to the results of Beran and Schützner ([5], Theorems 1–4) pertaining to the three-parametric LARCH model in (4.2) with $b_j = cj^{d-1}$, except that Beran and Schützner [5] deal with a modified QML estimation involving a "tuning parameter" $\epsilon > 0$. As explained by Beran and Schützner ([5], Section 3.2), the convergence rate of $\nabla\widetilde{L}_n(\theta_0)$ and $\widetilde{\theta}_n$ (based on nonstationary truncated observable series in (4.11)) is, apparently, too slow to guarantee asymptotic normality, this fact being a consequence of long memory in volatility and the main reason for introducing estimators $\widetilde{\theta}_n^{(\beta)}$ in (4.13). Theorems 4.1 and 4.2 are based on subsequent Lemmas 4.1–4.4 which describe properties of the likelihood processes defined in (4.8), (4.11) and (4.12). As noted in Section 4.1, our proofs use techniques different from those of Beran and Schützner [5], which rely on the explicit Volterra series representation of a stationary solution of the LARCH model.

For multi-index $\boldsymbol{i} = (i_1, ..., i_5) \in \mathbb{N}^5$, $\boldsymbol{i} \neq \boldsymbol{0} = (0, \cdots, 0)$, $|\boldsymbol{i}| := i_1 + \cdots + i_5$, denote partial derivative $\partial^{\boldsymbol{i}} := \partial^{|\boldsymbol{i}|}/\prod_{j=1}^5 \partial^{i_j}\theta_j$.

Lemma 4.1. *Let* $\mathrm{E}|r_t|^p < \infty$*, for some integer* $p \geq 1$*. Then for any* $\boldsymbol{i} \in \mathbb{N}^5$*,* $0 < |\boldsymbol{i}| \leq p$*,*

$$\mathrm{E} \sup_{\theta \in \Theta} \left| \partial^{\boldsymbol{i}} l_t(\theta) \right| < \infty. \tag{4.19}$$

Moreover, if $\mathrm{E}|r_t|^{p+\epsilon} < \infty$ *for some* $\epsilon > 0$ *and* $p \in \mathbb{N}$*, then for any* $\boldsymbol{i} \in \mathbb{N}^5$*,* $0 \leq |\boldsymbol{i}| \leq p$*,*

$$\mathrm{E} \sup_{\theta \in \Theta} \left| \partial^{\boldsymbol{i}} \left(l_t(\theta) - \widetilde{l}_t(\theta) \right) \right| \to 0, \qquad t \to \infty. \tag{4.20}$$

Lemma 4.2. *The function* $L(\theta), \theta \in \Theta$*, in (4.14) is bounded and continuous. Moreover, it attains its unique minimum at* $\theta = \theta_0$*.*

Lemma 4.3. *Let* $\mathrm{E}r_0^4 < \infty$*. Then matrices* $A(\theta)$ *and* $B(\theta)$ *in (4.15) are well-defined and strictly positive definite for any* $\theta \in \Theta$*.*

Write $|\cdot|$ for the Euclidean norm in $\mathbb{R}^5$ and in $\mathbb{R}^5 \otimes \mathbb{R}^5$ (the matrix norm).

Lemma 4.4. (i) *Let* $\mathrm{E}|r_t|^3 < \infty$*. Then*

$$\sup_{\theta \in \Theta} |L_n(\theta) - L(\theta)| \overset{a.s.}{\to} 0 \qquad \text{and} \qquad \mathrm{E} \sup_{\theta \in \Theta} |L_n(\theta) - \widetilde{L}_n(\theta)| \to 0. \tag{4.21}$$

(ii) *Let* $\mathrm{E}r_t^4 < \infty$*. Then* $\nabla L(\theta) = \mathrm{E}\nabla l_t(\theta)$ *and*

$$\sup_{\theta \in \Theta} |\nabla L_n(\theta) - \nabla L(\theta)| \overset{a.s.}{\to} 0 \qquad \text{and} \qquad \mathrm{E} \sup_{\theta \in \Theta} |\nabla L_n(\theta) - \nabla \widetilde{L}_n(\theta)| \to 0. \tag{4.22}$$

(iii) *Let* $\mathrm{E}|r_t|^5 < \infty$*. Then* $\nabla^T \nabla L(\theta) = \mathrm{E}\nabla^T \nabla \ell_t(\theta) = B(\theta)$ *(Equation (4.15))* *and*

$$\sup_{\theta\in\Theta} |\nabla^T \nabla L_n(\theta) - \nabla^T \nabla L(\theta)| \stackrel{a.s.}{\to} 0, \tag{4.23}$$

$$\mathrm{E} \sup_{\theta\in\Theta} |\nabla^T \nabla L_n(\theta) - \nabla^T \nabla \widetilde{L}_n(\theta)| \to 0. \tag{4.24}$$

Remark 4.3. As noted earlier, the moment conditions of Theorems 4.1 and 4.2 are similar to those of Beran and Schützner [5] for the LARCH model. Particularly, condition (M_5') of Beran and Schützner ([5], Theorems 2 and 5) for asymptotic normality of estimators ensures $\mathrm{E}|r_t|^5 < \infty$. This situation is very different from those of GARCH models where strong consistency and asymptotic normality of QML estimators hold under virtually no moment assumption on the observed process; see, for example, Francq and Zakoian ([28], Chapter 7). The main reason for this difference seems to be the fact that differentiation with respect to d of $Y_t(d) = \sum_{j=1}^{\infty} j^{d-1} r_{t-j}$ in (4.9) affects all terms of this series and results in "new" long memory processes $\partial^i Y_t(d)/\partial d^i = \sum_{j=1}^{\infty} j^{d-1} (\log j)^i r_{t-j}, i = 1, 2, 3,$ which are not bounded by $C|Y_t(d)|$ or $C\sigma_t^2(\theta)$. Therefore, derivatives of $\sigma_t^{-2}(\theta)$ in (4.9) are much more difficult to control than in the GARCH case, where these quantities are bounded; see Francq and Zakoian [28], proof of Theorem 7.2.

Remark 4.4. We expect that our results can be extended to more general parametric coefficients, for example, fractional filters $b_j(c, d), j \geq 1$, with the transfer function $\sum_{j=1}^{\infty} \mathrm{e}^{-\mathrm{i}jx} b_j(c, d) = g(c, d)((1-\mathrm{e}^{\mathrm{i}x})^{-d} - 1), x \in [-\pi, \pi]$, where $g(c, d)$ is a smooth function of $(c, d) \in (0, \infty) \times (0, 1/2)$. Particularly,

$$b_j(c, d) := g(c, d) \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)} \sim \frac{g(c, d)}{\Gamma(d)} j^{d-1}, \quad j \to \infty, \tag{4.25}$$

and $\sum_{j=1}^{\infty} b_j^2(c, d) = g^2(c, d)(\Gamma(1 - 2d) - \Gamma^2(1 - d))/\Gamma^2(1 - d)$; see, for example, Giraitis *et al.* ([36], Chapter 7). See also Beran and Schützner ([5], Section 2.2). An important condition used in our proofs and satisfied by $b_j(c,d)$ in (4.25) is that partial derivatives $\partial_d^i b_j(c,d), i = 1,2,3,$ decay at a similar rate j^{d-1} (modulus a slowly varying factor). Particularly, for ARFIMA$(0,d,0)$ coefficients $b_j^0(d) := \Gamma(j+d)/\Gamma(d)\Gamma(j+1) = \prod_{k=1}^j \frac{d+k-1}{k}$, it easily follows that

$$\partial_d b_j^0(d) = b_j^0(d) \sum_{k=1}^j \frac{1}{d+k-1} \sim b_j^0(d) \log j \sim \Gamma(d)^{-1} j^{d-1} \log j,$$

and, similarly,

$$\partial_d^i b_j^0(d) \sim b_j^0(d)(\log j)^i \sim \Gamma(d)^{-1} j^{d-1} (\log j)^i,\ j \to \infty,\ i = 2,3.$$

4.5 Simulation study

We present a short simulation study of the performance of the QML estimation for the GQARCH model in (4.4). The GQARCH model in (4.4) with i.i.d. standard normal innovations $\{\zeta_t\}$ was simulated for $-m+1 \le t \le m$ and two sample sizes $m = 1000$ and $m = 5000$, using the recurrent formula in (4.1) with zero initial condition $\sigma_{-m} = 0$. The numerical optimization procedure minimized the QML function:

$$\widetilde{L}_m = \frac{1}{m} \sum_{t=1}^m \left(\frac{r_t^2}{\sigma_t^2} + \log \sigma_t^2 \right), \tag{4.26}$$

with

$$r_t = \zeta_t \sigma_t, \quad \sigma_t^2 = \omega^2 + \left(a + c \sum_{j=1}^{t+m-1} j^{d-1} r_{t-j} \right)^2 + \gamma \sigma_{t-1}^2, \quad t = 1, \cdots, m. \tag{4.27}$$

The QML function in (4.26) can be viewed as a "realistic proxy" to the QML function $\widetilde{L}_n(\theta)$ in (4.13) with $m = n^\beta$ since (4.26) and (4.27) similar to (4.13) use "auxiliary" observations in addition to $r_1, \cdots, r_m$ for computation of m likelihoods in (4.26). However, the number of "auxiliary" observations in (4.26) equals m and does not grow as $m^{1/\beta} = n, 0 < \beta < 1 - 2d < 1$, in the case of (4.27) and Theorem 4.2 (ii), which is completely unrealistic. Despite the violation of the condition $m = n^\beta$ of Theorem 4.2 (ii) in our simulation study, differences between the sample root mean square errors (RMSEs) and the theoretical standard deviations are not vital (and sometimes even insignificant); see Table 4.5 below.

Finite-sample performance of the QML estimator $\widetilde{\theta}_m$ minimizing (4.26) was studied for fixed values of parameters $\gamma_0 = 0.7, a_0 = -0.2, c_0 = 0.2$, and different values of $\omega_0 = 0.1, 0.01$, and the long memory parameter $d_0 = 0.1, 0.2, 0.3, 0.4$. The aforementioned choice of $\theta_0 = (\gamma_0, \omega_0, a_0, d_0, c_0)$ can be explained by an observation that the QML estimation of γ_0, a_0, c_0 appears to be more accurate and stable in comparison with the estimation of ω_0 and d_0. The small values of ω_0 in our experiment reflect the fact that in most real data studied by us, the estimated QML value of ω_0 was smaller than 0.05.

The numerical QML minimization was performed using the MATLAB language for technical computing, under the following constraints:

$$0.005 \le \gamma \le 0.989, \quad 0 \le \omega \le 2, \quad -2 \le a \le 2, \quad 0 \le d \le 0.5,$$

and the value of c in the optimization procedure is chosen in such a way that would guarantee Assumption (B) (v) with appropriate $0 < c_i(d,\gamma), i = 1,2$.

The results of the simulation experiment are presented in Table 4.5, which shows the sample RMSEs of the QML estimates $\widetilde{\theta}_m = (\widetilde{\gamma}_m, \widetilde{\omega}_m, \widetilde{a}_m, \widetilde{d}_m, \widetilde{c}_m)$ with 100 independent replications, for two sample lengths $m = 1000$ and $m = 5000$ and the aforementioned choices of $(\gamma_0, \omega_0, a_0, d_0, c_0)$. The sample RMSEs in Table 4.5 are confronted with standard deviations (in parantheses) of the infinite past estimator in (4.7) computed according to Theorem 4.1 (ii) with $\Sigma(\theta_0)$ obtained by inverting a simulated matrix $B(\theta_0)/\kappa_4$.

A general impression from Table 4.5 is that theoretical standard deviations (bracketed entries) are generally smaller than the sample RMSEs; however, these differences become less pronounced with the increase of m and in some cases (e.g. when $\omega_0 = 0.1, m = 5000$) they seem to be insignificant. Some tendencies in Table 4.5 are quite surprising, particularly, the decrease of the theoretical standard deviations and most of sample RMSEs as d_0 increases. Also note a sharp increase of theoretical standard deviations of $\widehat{\omega}_n$ when $\omega_0 = 0.01$, which can be explained by the fact that the derivative $\partial_\omega \sigma_t^2(\theta_0) = 2\omega_0/(1-\gamma_0)$ becomes very small with ω_0, resulting in a small entry of $B(\theta_0)$ and a large entry of $\Sigma(\theta_0)$. On the other hand, the RMSEs in Table 4.5 appear to be more stable and less dependent on θ_0 compared with the bracketed entries (in particular this applies to errors of $\widetilde{\omega}_m$ and $\widetilde{d}_m$).

Table 4.1: Sample RMSE of the finite past QML estimates $\widetilde{\theta}_m$, received optimizing (4.26), of $\theta_0 = (\gamma_0, \omega_0, a_0, d_0, c_0)$ of the GQARCH process in (4.4) for $a_0 = -0.2, c_0 = 0.2, \gamma_0 = 0.7$ and different values of ω_0, d_0. The number of replications is 100. The quantities in parantheses stand for asymptotic standard deviations of the estimator $\widetilde{\theta}_n^{(\beta)}, n^\beta = m$ following Theorem 4.1 (ii).

		ω_0=0.1				
m	d_0	$\widetilde{\gamma}_m$	$\widetilde{\omega}_m$	$\widetilde{a}_m$	$\widetilde{d}_m$	$\widetilde{c}_m$
1000	0.1	0.076 (0.053)	0.046 (0.037)	0.032 (0.023)	0.090 (0.079)	0.027 (0.031)
	0.2	0.051 (0.048)	0.043 (0.027)	0.027 (0.020)	0.076 (0.060)	0.030 (0.027)
	0.3	0.069 (0.043)	0.033 (0.018)	0.026 (0.017)	0.063 (0.041)	0.030 (0.022)
	0.4	0.047 (0.039)	0.028 (0.013)	0.025 (0.015)	0.043 (0.029)	0.022 (0.019)
5000	0.1	0.023 (0.024)	0.018 (0.016)	0.011 (0.010)	0.035 (0.033)	0.014 (0.014)
	0.2	0.020 (0.021)	0.011 (0.011)	0.010 (0.009)	0.028 (0.021)	0.012 (0.012)
	0.3	0.019 (0.019)	0.010 (0.008)	0.010 (0.008)	0.020 (0.013)	0.010 (0.010)
	0.4	0.022 (0.017)	0.007 (0.005)	0.011 (0.007)	0.014 (0.009)	0.010 (0.008)
		ω_0=0.01				
m	d_0	$\widetilde{\gamma}_m$	$\widetilde{\omega}_m$	$\widetilde{a}_m$	$\widetilde{d}_m$	$\widetilde{c}_m$
1000	0.1	0.060 (0.046)	0.040 (0.296)	0.020 (0.019)	0.073 (0.071)	0.022 (0.029)
	0.2	0.044 (0.040)	0.035 (0.203)	0.020 (0.016)	0.073 (0.048)	0.022 (0.024)
	0.3	0.045 (0.033)	0.028 (0.117)	0.018 (0.012)	0.044 (0.029)	0.020 (0.019)
	0.4	0.040 (0.025)	0.038 (0.047)	0.024 (0.009)	0.034 (0.016)	0.020 (0.013)
5000	0.1	0.021 (0.020)	0.032 (0.125)	0.009 (0.008)	0.031 (0.028)	0.013 (0.013)
	0.2	0.018 (0.017)	0.024 (0.085)	0.007 (0.007)	0.020 (0.018)	0.010 (0.011)
	0.3	0.019 (0.015)	0.021 (0.046)	0.008 (0.006)	0.013 (0.011)	0.008 (0.009)
	0.4	0.016 (0.012)	0.013 (0.017)	0.007 (0.004)	0.011 (0.006)	0.009 (0.006)

4.6 Proofs

We use the following moment inequality by Burkholder [14] and Rosenthal [64].

Proposition 4.2. *Let $p \geq 2$ and $\{Y_j\}$ be a martingale difference sequence such that $\mathrm{E}|Y_j|^p < \infty$; $\mathrm{E}[Y_j|Y_1, \cdots, Y_{j-1}] = 0$, $j = 2, 3, \ldots$. Then there exists a* *constant K_p depending only on p and such that*

$$\mathrm{E}\Big|\sum_{j=1}^{\infty} Y_j\Big|^p \;\le\; K_p \left(\sum_{j=1}^{\infty} (\mathrm{E}|Y_j|^p)^{2/p}\right)^{p/2}. \tag{4.28}$$

Proof of Lemma 4.1. We use the following (Faà di Bruno) differentiation rule:

$$\begin{aligned}
\partial^{\boldsymbol{i}}\sigma_t^{-2}(\theta) &= \sum_{\nu=1}^{|\boldsymbol{i}|} \frac{(-1)^{\nu}\nu!}{\sigma_t^{2(1+\nu)}(\theta)} \sum_{\boldsymbol{j}_1+\cdots+\boldsymbol{j}_\nu=\boldsymbol{i}} \chi_{\boldsymbol{j}_1,\cdots,\boldsymbol{j}_\nu} \prod_{k=1}^{\nu} \partial^{\boldsymbol{j}_k}\sigma_t^2(\theta), \qquad (4.29)\\
\partial^{\boldsymbol{i}}\log\sigma_t^{2}(\theta) &= \sum_{\nu=1}^{|\boldsymbol{i}|} \frac{(-1)^{\nu-1}(\nu-1)!}{\sigma_t^{2\nu}(\theta)} \sum_{\boldsymbol{j}_1+\cdots+\boldsymbol{j}_\nu=\boldsymbol{i}} \chi_{\boldsymbol{j}_1,\cdots,\boldsymbol{j}_\nu} \prod_{k=1}^{\nu} \partial^{\boldsymbol{j}_k}\sigma_t^2(\theta),
\end{aligned}$$

where the sum $\sum_{\boldsymbol{j}_1+\cdots+\boldsymbol{j}_\nu=\boldsymbol{i}}$ is taken over decompositions of $\boldsymbol{i}$ into a sum of ν multi-indices $\boldsymbol{j}_k \neq \mathbf{0}, k = 1, \cdots, \nu$, and $\chi_{\boldsymbol{j}_1,\cdots,\boldsymbol{j}_\nu}$ is a combinatorial factor depending only on $\boldsymbol{j}_k, 1 \le k \le \nu$.

Let us prove (4.19). We have $\left|\partial^{\boldsymbol{i}} l_t(\theta)\right| \le r_t^2\left|\partial^{\boldsymbol{i}}\sigma_t^{-2}(\theta)\right| + \left|\partial^{\boldsymbol{i}}\log\sigma_t^2(\theta)\right|$. Hence using (4.29) and the fact that $\sigma_t^2(\theta) \ge \omega^2/(1-\gamma) \ge \omega_1^2/(1-\gamma_2) > 0$, we obtain

$$\sup_{\theta\in\Theta}\left|\partial^{\boldsymbol{i}} l_t(\theta)\right| \le C(r_t^2+1)\sum_{\nu=1}^{|\boldsymbol{i}|}\sum_{\boldsymbol{j}_1+\cdots+\boldsymbol{j}_\nu=\boldsymbol{i}}\prod_{k=1}^{\nu}\sup_{\theta\in\Theta}\left(\left|\partial^{\boldsymbol{j}_k}\sigma_t^2(\theta)\right|/\sigma_t(\theta)\right).$$

Therefore, by Hölder's inequality,

$$\begin{aligned}
&\mathrm{E}\sup_{\theta\in\Theta}\left|\partial^{\boldsymbol{i}} l_t(\theta)\right| \le C\left(\mathrm{E}(r_t^2+1)^{(2+p)/2}\right)^{2/(2+p)} \times\\
&\times\sum_{\nu=1}^{|\boldsymbol{i}|}\sum_{\boldsymbol{j}_1+\cdots+\boldsymbol{j}_\nu=\boldsymbol{i}}\prod_{k=1}^{\nu}\mathrm{E}^{1/q_k}\left(\sup_{\theta\in\Theta}\left|\partial^{\boldsymbol{j}_k}\sigma_t^2(\theta)\right|/\sigma_t(\theta)\right)^{q_k}, \qquad (4.30)
\end{aligned}$$

where $\sum_{j=1}^{\nu} 1/q_j \le p/(2+p)$. Note $|\boldsymbol{i}| = \sum_{k=1}^{\nu} |\boldsymbol{j}_k|$ and thus the choice $q_k = (2+p)/|\boldsymbol{j}_k|$ satisfies $\sum_{j=1}^{\nu} 1/q_j = \sum_{k=1}^{\nu} |\boldsymbol{j}_k|/(2+p) \le p/(2+p)$. Using (4.30) and condition $\mathrm{E}|r_t|^{2+p} \le C$, relation (4.19) follows from

$$\mathrm{E} \sup_{\theta\in\Theta} \Big(\big|\partial^{\boldsymbol{j}} \sigma_t^2(\theta)\big|/\sigma_t(\theta)\Big)^{(2+p)/|\boldsymbol{j}|} \ < \ \infty, \tag{4.31}$$

for any multi-index $\boldsymbol{j} \in \mathbb{N}^5$, $1 \le |\boldsymbol{j}| \le p$.

Consider first the case $|\boldsymbol{j}| = 1$, or the partial derivative $\partial_i \sigma_t^2(\theta) = \partial\sigma_t^2(\theta)/\partial\theta_i, 1 \le i \le 5$. We have

$$\partial_i \sigma_t^2(\theta) \ = \ \begin{cases} \sum_{\ell=1}^{\infty} \ell\gamma^{\ell-1}\big\{\omega^2 + \big(a + cY_{t-\ell}(d)\big)^2\big\}, & \theta_i = \gamma, \\ \sum_{\ell=0}^{\infty} \gamma^{\ell} 2\omega, & \theta_i = \omega, \\ \sum_{\ell=0}^{\infty} \gamma^{\ell} 2\big(a + cY_{t-\ell}(d)\big), & \theta_i = a, \\ \sum_{\ell=0}^{\infty} \gamma^{\ell} 2\big(a + cY_{t-\ell}(d)\big)Y_{t-\ell}(d), & \theta_i = c, \\ \sum_{\ell=0}^{\infty} \gamma^{\ell} 2c\big(a + cY_{t-\ell}(d)\big)\partial_d Y_{t-\ell}(d), & \theta_i = d. \end{cases} \tag{4.32}$$

We claim that there exist $C > 0, 0 < \bar\gamma < 1$, such that

$$\sup_{\theta\in\Theta} \left| \frac{\partial_i \sigma_t^2(\theta)}{\sigma_t(\theta)} \right| \ \le \ C(1 + J_{t,0} + J_{t,1}), \quad i = 1, \cdots, 5, \quad \text{where} \tag{4.33}$$

$$J_{t,0} \ := \ \sum_{\ell=0}^{\infty} \bar\gamma^{\ell} \sup_{d\in[d_1,d_2]} |Y_{t-\ell}(d)|, \qquad J_{t,1} \ := \ \sum_{\ell=0}^{\infty} \bar\gamma^{\ell} \sup_{d\in[d_1,d_2]} |\partial_d Y_{t-\ell}(d)|.$$

Consider (4.33) for $\theta_i = \gamma$. Using $\ell^2\gamma^{\ell-2} \le C\bar\gamma^{\ell}$ for all $\ell \ge 1, \gamma \in [\gamma_1, \gamma_2] \subset (0,1)$, and some $C > 0, 0 < \bar\gamma < 1$ together with Assumption (B) and Cauchy inequality, we obtain $|\partial_\gamma \sigma_t^2(\theta)|/\sigma_t(\theta) \le \big(\sum_{\ell=1}^{\infty} \ell^2\gamma^{\ell-2}\big\{\omega^2 + \big(a + cY_{t-\ell}(d)\big)^2\big)^{1/2} \le C(1 + J_{t,0})$ uniformly in $\theta \in \Theta$, proving (4.33) for $\theta_i = \gamma$. Similarly, $|\partial_c \sigma_t^2(\theta)|/\sigma_t(\theta) \le C(1 + J_{t,0})$ and $|\partial_d \sigma_t^2(\theta)|/\sigma_t(\theta) \le C(1 + J_{t,1})$.

Finally, for $\theta_i = \omega$ and $\theta_i = a$, (4.33) is immediate from (4.32), proving (4.33).

With (4.33) in mind, (4.31) for $|\boldsymbol{j}| = 1$ follows from

$$\mathrm{E}J_{t,i}^{2+p} \;=\; \mathrm{E}\big(\sum_{\ell=0}^{\infty} \bar{\gamma}^{\ell} \sup_{d\in[d_1,d_2]} |\partial_d^i Y_{t-\ell}(d)|\big)^{2+p} \;<\; \infty, \quad i = 0, 1. \tag{4.34}$$

Using Minkowski's inequality and stationarity of $\{Y_t(d)\}$, we obtain

$$\begin{aligned}\mathrm{E}^{1/(2+p)} J_{t,i}^{2+p} \;&\le\; \sum_{\ell=0}^{\infty} \bar{\gamma}^{\ell} \mathrm{E}^{1/(2+p)} \sup_d |\partial_d^i Y_{t-\ell}(d)|^{2+p} \\ &\le\; C\left(\mathrm{E}\sup_d |\partial_d^i Y_t(d)|^{2+p}\right)^{1/(2+p)},\end{aligned}$$

where

$$\partial_d^i Y_t(d) = \sum_{j=1}^{\infty} \partial_d^i j^{d-1} r_{t-j}.$$

Hence, using Beran and Schützner ([5]) Lemma 1 (b) and the inequality $xy \le x^q/q + y^{q'}/q', x, y > 0, 1/q + 1/q' = 1$, we obtain

$$\begin{aligned}\sum_{i=0}^{1} \mathrm{E}J_{t,i}^{2+p} \;&\le\; C\sum_{i=0}^{1} \mathrm{E} \sup_{d\in[d_1,d_2]} |\partial_d^i Y_t(d)|^{2+p} \\ &\le\; C\sum_{i=0}^{2} \sup_{d\in[d_1,d_2]} \mathrm{E}|\partial_d^i Y_t(d)|^{2+p} \;<\; \infty,\end{aligned} \tag{4.35}$$

since

$$\begin{aligned}\sup_{d\in[d_1,d_2]} \mathrm{E}|\partial_d^i Y_t(d)|^{2+p} \;&\le\; C \sup_{d\in[d_1,d_2]} \left(\sum_{j=1}^{\infty} \left(\partial_d^i j^{d-1}\right)^2 \left(\mathrm{E}|r_{t-j}|^{2+p}\right)^{2/(2+p)}\right)^{(2+p)/2} \\ &<\; \infty,\end{aligned}$$

according to condition $\mathrm{E}|r_t|^{2+p} \;<\; C$, Rosenthal's inequality in (4.28) and the fact that $\sup_{d\in[d_1,d_2]}\sum_{j=1}^{\infty}(\partial_d^i j^{d-1})^2 \le \sup_{d\in[d_1,d_2]}\sum_{j=1}^{\infty} j^{2(d-1)}(1+\log^2 j)^2 < C, i = 0,1,2$. This proves (4.31) for $|\boldsymbol{j}| = 1$.

The proof of (4.31) for $2 \le |\boldsymbol{j}| \le p$ is simpler since it reduces to

$$\mathrm{E}\sup_{\theta\in\Theta}\left|\partial^{\boldsymbol{j}}\sigma_t^2(\theta)\right|^{(p+2)/2} < \infty, \qquad 2 \le |\boldsymbol{j}| \le p. \tag{4.36}$$

Recall $\theta_1 = \gamma$ and $\boldsymbol{j}' := \boldsymbol{j} - (j_1,0,0,0,0) = (0,j_2,j_3,j_4,j_5)$. If $\boldsymbol{j}' = \mathbf{0}$, then $\sup_{\theta\in\Theta}\left|\partial^{\boldsymbol{j}}\sigma_t^2(\theta)\right| \le CJ_{t,0}$ follows as in (4.33) implying (4.36) as in (4.35) above. Next, let $\boldsymbol{j}' \neq \mathbf{0}$. Denote

$$Q_t^2(\theta) := \omega^2 + \big(a + cY_t(d)\big)^2, \tag{4.37}$$

so that $\sigma_t^2(\theta) = \sum_{\ell=0}^{\infty}\gamma^\ell Q_{t-\ell}^2(\theta)$. We have with $m := j_1 \ge 0$ that $|\partial^{\boldsymbol{j}}\sigma_t^2(\theta)| \le \sum_{\ell=m}^{\infty}(\ell!/(\ell-m)!)\gamma^{\ell-m}\,|\partial^{\boldsymbol{j}'}Q_{t-\ell}^2(\theta)|$ and (4.31) follows from

$$\mathrm{E}\sup_{\theta\in\Theta}\left|\partial^{\boldsymbol{j}}Q_t^2(\theta)\right|^{(p+2)/2} < \infty. \tag{4.38}$$

For $j_2 \neq 0$ (recall $\theta_2 = \omega$) the derivative in (4.38) is trivial so that it suffices to check (4.38) only for $j_1 = 0$. Then applying Faà di Bruno's rule we get

$$\left|\partial^{\boldsymbol{j}}Q_t^2(\theta)\right|^{(p+2)/2} \le C\sum_{\boldsymbol{j}_1+\boldsymbol{j}_2=\boldsymbol{j}}\left|\partial^{\boldsymbol{j}_1}(a+cY_t(d))\right|^{(p+2)/2}\left|\partial^{\boldsymbol{j}_2}(a+cY_t(d))\right|^{(p+2)/2},$$

and hence, (4.38) reduces to

$$\mathrm{E}\sup_{\theta\in\Theta}\left|\partial^{\boldsymbol{j}}(a+cY_t(d))\right|^{p+2} < \infty, \qquad 0 \le |\boldsymbol{j}| \le p,$$

whose proof is similar to (4.34) above. This ends the proof of (4.19).

The proof of (4.20) is similar. We have $|\partial^{\boldsymbol{i}}(l_t(\theta) - \widetilde{l}_t(\theta))| \le r_t^2|\partial^{\boldsymbol{i}}(\sigma_t^{-2}(\theta) -$

$\widetilde{\sigma}_t^{-2}(\theta))| + |\partial^{\boldsymbol{i}}(\log \sigma_t^2(\theta) - \log \widetilde{\sigma}_t^2(\theta))|$. Hence, using Hölder's inequality similarly as in the proof of (4.19), it suffices to show

$$\mathrm{E} \sup_{\theta \in \Theta} \left| \partial^{\boldsymbol{i}} (\sigma_t^{-2}(\theta) - \widetilde{\sigma}_t^{-2}(\theta)) \right|^{\frac{p+2}{p}} \to 0 \tag{4.39}$$

and

$$\mathrm{E} \sup_{\theta \in \Theta} \left| \partial^{\boldsymbol{i}} (\log \sigma_t^2(\theta) - \log \widetilde{\sigma}_t^2(\theta)) \right|^{\frac{p+2}{p}} \to 0. \tag{4.40}$$

Below, we prove only the relation (4.39), the proof of (4.40) being analogous.

Using the differentiation rule in (4.29), we have that

$$\left| \partial^{\boldsymbol{i}} (\sigma_t^{-2}(\theta) - \widetilde{\sigma}_t^{-2}(\theta)) \right| \;\leq\; C \sum_{\nu=1}^{|\boldsymbol{i}|} \sum_{\boldsymbol{j}_1 + \cdots + \boldsymbol{j}_\nu = \boldsymbol{i}} \left| W_t^{\boldsymbol{j}_1, \cdots, \boldsymbol{j}_\nu}(\theta) - \widetilde{W}_t^{\boldsymbol{j}_1, \cdots, \boldsymbol{j}_\nu}(\theta) \right|,$$

where

$$\begin{aligned} W_t^{\boldsymbol{j}_1, \cdots, \boldsymbol{j}_\nu}(\theta) &:= \sigma_t^{-2(1+\nu)}(\theta) \textstyle\prod_{k=1}^{\nu} \partial^{\boldsymbol{j}_k} \sigma_t^2(\theta), \\ \widetilde{W}_t^{\boldsymbol{j}_1, \cdots, \boldsymbol{j}_\nu}(\theta) &:= \widetilde{\sigma}_t^{-2(1+\nu)}(\theta) \textstyle\prod_{k=1}^{\nu} \partial^{\boldsymbol{j}_k} \widetilde{\sigma}_t^2(\theta). \end{aligned}$$

Whence, (4.39) follows from

$$\sup_{\theta \in \Theta} \left| W_t^{\boldsymbol{j}_1, \cdots, \boldsymbol{j}_\nu}(\theta) - \widetilde{W}_t^{\boldsymbol{j}_1, \cdots, \boldsymbol{j}_\nu}(\theta) \right| \overset{\mathrm{P}}{\to} 0, \qquad t \to \infty \tag{4.41}$$

and

$$\mathrm{E} \sup_{\theta \in \Theta} \left(\left| W_t^{\boldsymbol{j}_1, \cdots, \boldsymbol{j}_\nu}(\theta) \right| + \left| \widetilde{W}_t^{\boldsymbol{j}_1, \cdots, \boldsymbol{j}_\nu}(\theta) \right| \right)^{(p+2+\epsilon)/p} \;\leq\; C < \infty, \tag{4.42}$$

for some constants $\epsilon > 0$ and $C > 0$ independent of t. In turn, (4.41) and (4.42) follow from

$$\sup_{\theta\in\Theta} \left|\partial^{\boldsymbol{j}} (\sigma_t^2(\theta) - \widetilde{\sigma}_t^2(\theta))\right| \stackrel{\mathrm{P}}{\to} 0, \qquad t \to \infty \tag{4.43}$$

and

$$\begin{aligned} \mathrm{E}\sup_{\theta\in\Theta} \Big(\left|\partial^{\boldsymbol{j}} \sigma_t^2(\theta)\right|/\sigma_t(\theta)\Big)^{(2+p+\epsilon)/|\boldsymbol{j}|} &< C, \qquad (4.44)\\ \mathrm{E}\sup_{\theta\in\Theta} \Big(\left|\partial^{\boldsymbol{j}} \widetilde{\sigma}_t^2(\theta)\right|/\widetilde{\sigma}_t(\theta)\Big)^{(2+p+\epsilon)/|\boldsymbol{j}|} &< C, \end{aligned}$$

for any multi-index $\boldsymbol{j}$ such that $|\boldsymbol{j}| \geq 0$ and $1 \leq |\boldsymbol{j}| \leq p$, respectively.

Using condition $\mathrm{E}|r_t|^{2+p+\epsilon} < C$, relations in (4.44) can be proved analogously to (4.31), and we omit the details. Consider (4.43). Split $\sigma_t^2(\theta) - \widetilde{\sigma}_t^2(\theta) = U_{t,1}(\theta) + U_{t,2}(\theta)$, where

$$\begin{aligned} U_{t,1}(\theta) &:= \sum_{\ell=1}^{t-1} \gamma^\ell \left\{ \big(a + cY_{t-\ell}(d)\big)^2 - \big(a + c\widetilde{Y}_{t-\ell}(d)\big)^2 \right\}, \qquad (4.45)\\ U_{t,2}(\theta) &:= \sum_{\ell=t}^{\infty} \gamma^\ell \left\{ \omega^2 + \big(a + cY_{t-\ell}(d)\big)^2 \right\}. \end{aligned}$$

Then $\sup_{\theta\in\Theta} |\partial^{\boldsymbol{j}} U_{t,i}(\theta)| \stackrel{\mathrm{P}}{\to} 0, t \to \infty, i = 1, 2$, follows by using Assumption (B) and considering the bounds on the derivatives as in the proof of (4.31). For instance, let us prove (4.43) for $\partial^{\boldsymbol{j}} = \partial_d, |\boldsymbol{j}| = 1$. We have

$$\begin{aligned} \left|\partial_d U_{t,1}(\theta)\right| &\leq C \sum_{\ell=1}^{t-1} \gamma^\ell \left\{ \Big(1 + \left|\bar{Y}_{t-\ell}(d)\right|\Big) \left|\partial_d \Big(Y_{t-\ell}(d) - \widetilde{Y}_{t-\ell}(d)\Big)\right| + \right.\\ &+ \left. \left|\partial_d Y_{t-\ell}(d)\right| \left|Y_{t-\ell}(d) - \widetilde{Y}_{t-\ell}(d)\right| \right\}. \end{aligned}$$

Hence, $\sup_{\theta\in\Theta} |\partial_d U_{t,1}(\theta)| \xrightarrow{\text{P}} 0$ follows from $0 \le \gamma \le \gamma_2 < 1$ and

$$\mathrm{E} \sup_{d\in[d_1,d_2]} \left(\left| Y_t(d) - \widetilde{Y}_t(d) \right|^2 + \left| \partial_d (Y_t(d) - \widetilde{Y}_t(d)) \right|^2 \right) \;\to\; 0, \tag{4.46}$$

$$\mathrm{E} \sup_{d\in[d_1,d_2]} \left(\left| Y_t(d) \right|^2 + \left| \widetilde{Y}_t(d) \right|^2 + \left| \partial_d Y_t(d) \right|^2 + \left| \partial_d \widetilde{Y}_t(d) \right|^2 \right) \;\le\; C. \tag{4.47}$$

The proof of (4.47) mimics that of (4.35) and, therefore, is omitted. To show (4.46), note $Y_t(d) - \widetilde{Y}_t(d) = \sum_{j=t}^{\infty} j^{d-1} r_{t-j}$ and use a similar argument as in (4.35) to show that the l.h.s. of (4.47) does not exceed

$$C \sup_{d\in[d_1,d_2]} \sum_{i=0}^{2} \mathrm{E} \left| \partial_d^i (Y_t(d) - \widetilde{Y}_t(d)) \right|^2 \le C \sup_{d\in[d_1,d_2]} \sum_{j=t}^{\infty} j^{2(d-1)} \left(1 + \log^2 j \right) \to 0,$$

as $t \to \infty$. This proves (4.43) for $|\boldsymbol{j}| = 1$ and $\partial^{\boldsymbol{j}} = \partial_d$. The remaining cases in (4.43) follow similarly, and we omit the details. This proves (4.20) and completes the proof of Lemma 4.1. □

Proof of Lemma 4.2. We have

$$|L(\theta_1) - L(\theta_2)| \le \mathrm{E}|l_t(\theta_1) - l_t(\theta_2)| \le C\mathrm{E}|\sigma_t^2(\theta_1) - \sigma_t^2(\theta_2)|,$$

where the last expectation can be easily shown to vanish as $|\theta_1 - \theta_2| \to 0$, $\theta_1, \theta_2 \in \Theta$. This proves the first statement of the lemma. To show the second statement of the lemma, write

$$L(\theta) - L(\theta_0) = \mathrm{E} \left[\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} - \log \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} - 1 \right].$$

The function $f(x) := x - 1 - \log x > 0$ for $x > 0, x \ne 1$ and $f(x) = 0$ if and only if $x = 1$. Therefore, $L(\theta) \ge L(\theta_0), \forall\, \theta \in \Theta$, while $L(\theta) = L(\theta_0)$ is equivalent to

$$\sigma_t^2(\theta) = \sigma_t^2(\theta_0) \qquad (\mathrm{P}_{\theta_0} - \text{a.s.}). \tag{4.48}$$

Thus, it remains to show that (4.48) implies $\theta = \theta_0 = (\gamma_0, \omega_0, a_0, d_0, c_0)$. Consider the "projection" $P_s\xi = \mathrm{E}[\xi|\mathcal{F}_s] - \mathrm{E}[\xi|\mathcal{F}_{s-1}]$ of r.v. ξ, $\mathrm{E}|\xi| < \infty$, where $\mathcal{F}_s = \sigma(\zeta_u, u \le s)$ (see Section 4.2). Equation (4.48) implies

$$0 = P_s(\sigma_t^2(\theta) - \sigma_t^2(\theta_0)) = P_s(Q_t^2(\theta) - Q_t^2(\theta_0)) + (\gamma - \gamma_0) P_s \sigma_{t-1}^2(\theta_0), \quad \forall s \le t-1, \tag{4.49}$$

where $Q_t^2(\theta) = \omega^2 + \big(a + \sum_{u<t} b_{t-u}(\theta) r_u\big)^2$ is the same as in (4.37). We have

$$\begin{aligned} P_s Q_t^2(\theta) &= 2ab_{t-s}(\theta) r_s + 2b_{t-s}(\theta) r_s \sum_{u<s} b_{t-u}(\theta) r_u + \sum_{s \le u < t} b_{t-u}^2(\theta) P_s r_u^2 \\ &= 2ab_{t-s}(\theta)\zeta_s \sigma_s(\theta_0) + 2b_{t-s}(\theta)\zeta_s\sigma_s(\theta_0) \sum_{u<s} b_{t-u}(\theta) r_u + \\ &+ \sum_{s<u<t} b_{t-u}^2(\theta) P_s \sigma_u^2(\theta_0) + b_{t-s}^2(\theta)(\zeta_s^2 - 1)\sigma_s^2(\theta_0). \end{aligned} \tag{4.50}$$

Whence and from (4.49) for $s = t-1$ using $P_{t-1}\sigma_{t-1}^2(\theta_0) = 0$, we obtain

$$C_1(\theta, \theta_0)\zeta_{t-1}^2 + 2C_2(\theta, \theta_0)\zeta_{t-1} - C_1(\theta, \theta_0) = 0, \tag{4.51}$$

where

$$\begin{aligned} C_1(\theta, \theta_0) &:= \left(c^2 - c_0^2\right) \sigma_{t-1}(\theta_0), \\ C_2(\theta, \theta_0) &:= (ac - a_0 c_0) + \sum_{u<t-1} \left(c^2 (t-u)^{d-1} - c_0^2 (t-u)^{d_0 - 1}\right) r_u. \end{aligned}$$

Since $C_i(\theta, \theta_0), i = 1, 2,$ are $\mathcal{F}_{t-2}$-measurable, (4.51) implies

$$C_1(\theta, \theta_0) = C_2(\theta, \theta_0) = 0,$$

particularly, $c = c_0$ since $\sigma_{t-1}(\theta_0) \geq \omega > 0$. Then

$$0 = C_2(\theta, \theta_0) = c_0(a - a_0) + c_0^2 \sum_{u<t-1} ((t-u)^{d-1} - (t-u)^{d_0-1}) r_u,$$

and $\mathrm{E} r_u = 0$ lead to $a = a_0$ and next to

$$0 = \mathrm{E}\left(\sum_{u<t-1} \left((t-u)^{d-1} - (t-u)^{d_0-1}\right) r_u \right)^2 = \mathrm{E} r_0^2 \sum_{j\geq 2} \left(j^{d-1} - j^{d_0-1}\right)^2 = 0,$$

or $d = d_0$. Consequently, $P_s(Q_t^2(\theta) - Q_t^2(\theta_0)) = 0$ for any $s \leq t-1$ and hence $\gamma = \gamma_0$ in view of (4.49). Finally, $\omega = \omega_0$ follows from $\mathrm{E}\sigma_t^2(\theta) = \mathrm{E}\sigma_t^2(\theta_0)$ and the fact that $\omega > 0, \omega_0 > 0$. This proves $\theta = \theta_0$ as well as the lemm. □

Proof of Lemma 4.3. From (4.16), it suffices to show that

$$\nabla \sigma_t^2(\theta) \lambda^T = 0, \tag{4.52}$$

for some $\theta \in \Theta$ and $\lambda \in \mathbb{R}^5, \lambda \neq 0$ leads to a contradiction. To this end, we use a similar projection argument as in the proof of Lemma 4.2. First, note that $\sigma_t^2(\theta) = Q_t^2(\theta) + \gamma \sigma_{t-1}^2(\theta)$ implies

$$\nabla \sigma_t^2(\theta) = \left(0, \nabla_4 Q_t^2(\theta)\right) + \gamma \nabla \sigma_{t-1}^2(\theta) + (\nabla \gamma) \sigma_{t-1}^2(\theta),$$

where $\nabla_4 = (\partial/\theta_2, \cdots, \partial\theta_5)$. Hence and using the fact that (4.52) holds for any $t \in \mathbb{Z}$ by stationarity, from (4.52) we obtain

$$\left(\sigma_{t-1}^2(\theta), \nabla_4 Q_t^2(\theta)\right) \lambda^T = 0. \tag{4.53}$$

Thus,

$$(P_s \sigma_{t-1}^2(\theta), P_s \nabla_4^T Q_t^2(\theta))\lambda \ = \ 0, \qquad \forall\, s \leq t-1;$$

compare with (4.49). For $s = t-1$ using $P_{t-1}\sigma^2_{t-1}(\theta) = 0,\ P_{t-1}\nabla_4 Q^2_t(\theta) = \nabla_4 P_{t-1} Q^2_t(\theta)$ by differentiating (4.50) similarly to (4.51), we obtain

$$D_1(\lambda)\zeta^2_{t-1} + 2D_2(\lambda)\zeta_{t-1} - D_1(\lambda) = 0, \tag{4.54}$$

where $D_1(\lambda) := 2\lambda_5\sigma_{t-1}(\theta)$ and

$$\begin{aligned} D_2(\lambda) &:= \lambda_3 c + \lambda_5 a + 2\lambda_5 c \sum_{u<t-1} (t-u)^{d-1} r_u + \\ &+ \lambda_4 c^2 \sum_{u<t-1} (t-u)^{d-2} \log(t-u) r_u, \end{aligned}$$

$\lambda = (\lambda_1, \cdots, \lambda_5)^T$. As in (4.51), $D_i(\lambda), i = 1, 2$ are $\mathcal{F}_{t-2}$-measurable, (4.54) implying $D_i(\lambda) = 0, i = 1, 2$. Hence, $\lambda_5 = 0$ and then $D_2(\lambda) = 0$ reduces to $\lambda_3 c + \lambda_4 c^2 \sum_{u<t-1}(t-u)^{d-2}\log(t-u)r_u = 0$. By taking expectation and using $c \neq 0$, we obtain $\lambda_3 = 0$ and then $\lambda_4 = 0$ since $\mathrm{E}\big(\sum_{u<t-1}(t-u)^{d-2}\log(t-u)r_u\big)^2 \neq 0$. The aforementioned facts allow rewriting (4.53) as $2\omega\lambda_2 + \lambda_1\sigma^2_{t-1}(\theta) = 0$. Unless both λ_1 and λ_2 vanish, the last equation means that either $\lambda_1 \neq 0$ and $\{\sigma^2_t(\theta)\}$ is a deterministic process, which contradicts $c \neq 0$, or $\lambda_1 = 0, \lambda_2 \neq 0$ and $\omega = 0$, which contradicts $\omega \neq 0$. Lemma 4.3 is proved. □

Proof of Lemma 4.4. Consider the first relation in (4.21). The pointwise convergence $L_n(\theta) \overset{a.s.}{\to} L(\theta)$ follows by ergodicity of $\{l_t(\theta)\}$ and the uniform convergence in (4.21) from $\mathrm{E}\sup_{\theta\in\Theta}|\nabla l_t(\theta)| < \infty$ (cf. Beran and Schützner [5], proof of Lemma 3), which, in turn, follows from Lemma 4.1 (4.19) with $p = 1$. The proof of the second relation in (4.21) is immediate from Lemma 4.1 (4.20) with $p = 0, \epsilon = 1$. The proof of statements (ii) and (iii) using Lemma 4.1 is similar and is omitted. □

Proof of Theorem 4.1. (i) Follows from Lemmas 4.2 and 4.4 (i) using standard argument.

(ii) By Taylor's expansion,

$$0 \;=\; \nabla L_n(\widehat{\theta}_n) = \nabla L_n(\theta_0) + \nabla^T \nabla L_n(\theta_n^*)(\widehat{\theta}_n - \theta_0),$$

where $\theta_n^* \xrightarrow{\mathrm{P}} \theta_0$ since $\widehat{\theta}_n \xrightarrow{\mathrm{P}} \theta_0$. Then $\nabla^T \nabla L_n(\theta_n^*) \xrightarrow{\mathrm{P}} \nabla^T \nabla L(\theta_0)$ by Lemma 4.4 (4.23). Next, since $\{r_t^2/\sigma_t^2(\theta_0) - 1, \mathcal{F}_t, t \in \mathbb{Z}\}$ is a square-integrable and ergodic martingale difference sequence, the convergence $n^{1/2} \nabla L_n(\theta_0) \xrightarrow{d} N(0, A(\theta_0))$ follows by the martingale central limit theorem of Billingsley ([9], Theorem 23.1). Then (4.17) follows by Slutsky's theorem and (4.15).
□

Proof of Theorem 4.2. Part (i) follows from Lemmas 4.2 and 4.4 (i) as in the case of Theorem 4.1 (i). To prove part (ii), by Taylor's expansion,

$$0 \;=\; \nabla \widetilde{L}_n^{(\beta)}\big(\widetilde{\theta}_n^{(\beta)}\big) = \nabla \widetilde{L}_n^{(\beta)}(\theta_0) + \nabla^T \nabla \widetilde{L}_n^{(\beta)}\big(\widetilde{\theta}_n^*\big)\big(\widetilde{\theta}_n^{(\beta)} - \theta_0\big),$$

where $\widetilde{\theta}_n^* \xrightarrow{\mathrm{P}} \theta_0$ since $\widetilde{\theta}_n^{(\beta)} \xrightarrow{\mathrm{P}} \theta_0$. Then $\nabla^T \nabla \widetilde{L}_n^{(\beta)}(\theta_n^*) \to_{\mathrm{p}} \nabla^T \nabla L(\theta_0)$ by Lemma 4.4 (4.23)–(4.24). From the proof of Theorem 4.1 (ii), we have that

$$n^{\beta/2} \nabla L_n^{(\beta)}(\theta_0) \xrightarrow{d} N(0, A(\theta_0)),$$

where $L_n^{(\beta)}(\theta) := \frac{1}{[n^\beta]} \sum_{t=n-[n^\beta]+1}^{n} l_t(\theta)$. Hence, the central limit theorem in (4.18) follows from

$$I_n(\beta) := \mathrm{E}\big|\nabla \widetilde{L}_n^{(\beta)}(\theta_0) - \nabla L_n^{(\beta)}(\theta_0)\big| \;=\; o(n^{-\beta/2}). \tag{4.55}$$

We have $I_n(\beta) \le \sup_{n-[n^\beta] \le t \le n} \mathrm{E}\big|\nabla l_t(\theta_0) - \nabla \widetilde{l}_t(\theta_0)\big|$ and (4.55) follows from

$$\mathrm{E}\big|\nabla l_t(\theta_0) - \nabla\widetilde{l}_t(\theta_0)\big| \;=\; o(t^{-\beta/2}), \qquad t\to\infty. \tag{4.56}$$

Write $\|\xi\|_p := \mathrm{E}^{1/p}|\xi|^p$ for L^p-norm of r.v. ξ. Using

$$\big|\nabla(l_t(\theta_0)-\widetilde{l}_t(\theta_0))\big| \le r_t^2\big|\nabla(\sigma_t^{-2}(\theta_0)-\widetilde{\sigma}_t^{-2}(\theta_0))\big|+\big|\nabla(\log\sigma_t^2(\theta_0)-\log\widetilde{\sigma}_t^2(\theta_0))\big|,$$

and assumption $\mathrm{E}|r_t|^5<\infty$, relation (4.56) follows from

$$\begin{aligned}\|\sigma_t^{-4}\partial_i\sigma_t^2-\widetilde{\sigma}_t^{-4}\partial_i\widetilde{\sigma}_t^2\|_{5/3} &= O\big(t^{d_0-1/2}\log t\big) \quad \text{and} \\ \|\sigma_t^{-2}\partial_i\sigma_t^2-\widetilde{\sigma}_t^{-2}\partial_i\widetilde{\sigma}_t^2\|_{1} &= O\big(t^{d_0-1/2}\log t\big), \qquad i=1,\cdots,5,\end{aligned} \tag{4.57}$$

where $\sigma_t^2 := \sigma_t^2(\theta_0)$, $\widetilde{\sigma}_t^2 := \widetilde{\sigma}_t^2(\theta_0)$, $\partial_i\sigma_t^2 := \partial_i\sigma_t^2(\theta_0)$ and $\partial_i\widetilde{\sigma}_t^2 := \partial_i\widetilde{\sigma}_t^2(\theta_0)$. Subsequently, we prove only the first relation in (4.57), the proof of the second one being similar. We have

$$\sigma_t^{-4}\partial_i\sigma_t^2-\widetilde{\sigma}_t^{-4}\partial_i\widetilde{\sigma}_t^2 = \sigma_t^{-4}\widetilde{\sigma}_t^{-4}(\widetilde{\sigma}_t^2+\sigma_t^2)(\widetilde{\sigma}_t^2-\sigma_t^2)\partial_i\sigma_t^2+\widetilde{\sigma}_t^{-4}(\partial_i\sigma^2-\partial_i\widetilde{\sigma}_t^2).$$

Then using $\sigma_t^2\ge\omega_1^2/(1-\gamma_2)>0, \widetilde{\sigma}_t^2\ge\omega_1^2/(1-\gamma_2)>$, the first relation in (4.57) follows from

$$\|(\sigma_t^2-\widetilde{\sigma}_t^2)(\partial_i\sigma_t^2/\sigma_t)\|_{5/3} = O\big(t^{d_0-1/2}\big) \quad \text{and} \tag{4.58}$$

$$\|\partial_i\sigma_t^2-\partial_i\widetilde{\sigma}_t^2\|_{5/3} = O\big(t^{d_0-1/2}\log t\big), \qquad i=1,\cdots,5. \tag{4.59}$$

Consider (4.58). By Hölder's inequality,

$$\|(\sigma_t^2-\widetilde{\sigma}_t^2)(\partial_i\sigma_t^2/\sigma_t)\|_{5/3} \le \|\sigma_t^2-\widetilde{\sigma}_t^2\|_{5/2}\|\partial_i\sigma_t^2/\sigma_t\|_5,$$

where $\|\partial_i \sigma_t^2/\sigma_t\|_5 < C$ according to (4.31). Hence, (4.58) follows from

$$\|\sigma_t^2 - \widetilde{\sigma}_t^2\|_{5/2} \;=\; O\big(t^{d_0-1/2}\big). \tag{4.60}$$

To show (4.60), similarly as in the proof of (4.43), split $\sigma_t^2 - \widetilde{\sigma}_t^2 = U_{t,1} + U_{t,2}$, where $U_{t,i} := U_{t,i}(\theta_0), i = 1, 2$, are defined in (4.45), that is,

$$\begin{aligned} U_{t,1} &= \sum_{\ell=1}^{t-1} \gamma_0^\ell \left\{ \big(a_0 + c_0 Y_{t-\ell}\big)^2 - \big(a_0 + c_0 \widetilde{Y}_{t-\ell}\big)^2 \right\}, \\ U_{t,2} &= \sum_{\ell=t}^{\infty} \gamma_0^\ell \left\{ \omega_0^2 + \big(a_0 + c_0 Y_{t-\ell}\big)^2 \right\}, \end{aligned}$$

and $Y_t := Y_t(d_0), \widetilde{Y}_t := \widetilde{Y}_t(d_0)$. We have

$$\begin{aligned} |U_{t,1}| &\le C \sum_{\ell=1}^{t-1} \gamma_0^\ell |Y_{t-\ell} - \widetilde{Y}_{t-\ell}| \big(1 + |Y_{t-\ell}| + |\widetilde{Y}_{t-\ell}|\big), \\ |U_{t,2}| &\le C \sum_{\ell=t}^{\infty} \gamma_0^\ell \big(1 + |Y_{t-\ell}|^2\big), \end{aligned}$$

and hence

$$\begin{aligned} \|\sigma_t^2 - \widetilde{\sigma}_t^2\|_{5/2} &\le C \left\{ \sum_{\ell=1}^{t-1} \gamma_0^\ell \|(Y_{t-\ell} - \widetilde{Y}_{t-\ell})(1 + |Y_{t-\ell}| + |\widetilde{Y}_{t-\ell}|)\|_{5/2} + \right. \\ &\qquad \left. + \sum_{\ell=t}^{\infty} \gamma_0^\ell (1 + \|Y_{t-\ell}\|_5) \right\} \\ &\le C \left\{ \sum_{\ell=1}^{t-1} \gamma_0^\ell \|Y_{t-\ell} - \widetilde{Y}_{t-\ell}\|_5 + \sum_{\ell=t}^{\infty} \gamma_0^\ell \right\}, \end{aligned} \tag{4.61}$$

where we used the fact that $\|Y_t\|_5 < C$, $\|\widetilde{Y}_t\|_5 < C$ by $\|r_t\|_5 < C$ and Rosenthal's inequality in (4.28). In a similar way from (4.28) it follows that

$$\|Y_{t-\ell} - \widetilde{Y}_{t-\ell}\|_5 \le C\left\{\sum_{j>t-\ell} j^{2(d_0-1)}\right\}^{1/2} \le C(t-\ell)^{d_0-1/2}. \tag{4.62}$$

Substituting (4.62) with (4.61) we obtain

$$\|\sigma_t^2 - \widetilde{\sigma}_t^2\|_{5/2} \le C\left\{\sum_{\ell=1}^{t-1} \gamma_0^\ell (t-\ell)^{d_0-1/2} + \sum_{\ell=t}^{\infty} \gamma_0^\ell\right\} = O(t^{d_0-1/2}),$$

proving (4.60).

It remains to show (4.59). Similarly to the previous discussion, $\partial_i\sigma_t^2 - \partial_i\widetilde{\sigma}_t^2 = \partial_i U_{t,1} + \partial_i U_{t,2}$, where $\partial_i U_{t,j} := \partial_i U_{t,j}(\theta_0), j = 1,2$. Then (4.59) follows from

$$\|\partial_i U_{t,1}\|_{5/3} = O(t^{d_0-1/2}\log t) \quad \text{and} \quad \|\partial_i U_{t,2}\|_{5/3} = o(t^{d_0-1/2}), \quad i = 1,\cdots,5. \tag{4.63}$$

For $i = 1$, the proof of (4.63) is similar to (4.61). Consider (4.63) for $2 \le i \le 5$. Denote $V_t(\theta) := 2a + c(Y_t(d) + \widetilde{Y}_t(d))$, $V_t := V_t(\theta_0), \partial_i V_t := \partial_i V_t(\theta_0)$, and then

$$\|\partial_i U_{t,1}\|_{5/3} \le C\sum_{\ell=1}^{t-1} \gamma_0^\ell \Big\{\|\partial_i(Y_{t-\ell} - \widetilde{Y}_{t-\ell})\|_5 \|V_t\|_5 + \|Y_{t-\ell} - \widetilde{Y}_{t-\ell}\|_5 \|\partial_i V_t\|_5\Big\},$$

where $\partial_i(Y_{t-\ell} - \widetilde{Y}_{t-\ell}) = 0, \partial_i \ne \partial_d$ and

$$\begin{aligned}\|\partial_d(Y_t - \widetilde{Y}_t)\|_5 &= \|\sum_{j>t} j^{d_0-1}(\log j) r_{t-j}\|_5 \\ &\le C\Big\{\sum_{j>t} j^{2(d_0-1)}\log^2 j\Big\}^{1/2} = O(t^{d_0-1/2}\log t)\end{aligned}$$

similarly as in (4.62). Hence, the first relation in (4.63) follows from (4.62) and $\|\partial_i V_t\|_5 \le C(1 + \|\partial_d Y_{t-\ell}\|_5 + \|\partial_d \widetilde{Y}_{t-\ell}\|_5) \le C < \infty$ as in the proof of (4.59), and the proof of the second relation in (4.63) is analogous. This proves (4.56) and completes the proof of Theorem 4.2. □

4.7 Conclusion

In this chapter we studied the five-parametric QML estimation for a quadratic ARCH process with long memory and strictly positive conditional variance introduced by Doukhan *et al.* [22] and Grublytė and Škarnulis [40], which extends the QARCH model of Sentana [66] and the Linear ARCH (LARCH) model of Robinson [62]. Following Beran and Schützner [5] who studied a similar problem for the LARCH model, we discussed several QML estimators of unknown parameter $\theta_0 \in \mathbb{R}^5$ of our model, in particular, an estimator $\widehat{\theta}_n$ depending on observations $r_s, -\infty < s \le n$, from the infinite past, and a class of estimators $\widetilde{\theta}_n^{(\beta)}, 0 < \beta < 1$, only depending on $r_s, 1 \le s \le n$, and minimizing the sum of the last $m = n^\beta = o(n)$ likelihoods. Under assumptions similar to those of Beran and Schützner [5], we proved consistency and asymptotic normality of these estimators with the convergence rate $m^{1/2} = o(n^{(1-d_0)/2})$, where $d_0 \in (0, 1/2)$ is the long memory parameter. However, using estimator $\widetilde{\theta}_n^{(\beta)}$ is unrealistic because of the poor and *a priori* unknown convergence rate. In the simulation experiment, we studied the empirical performance of a realistic version of this estimator based on $m = n/2$ last likelihoods, for $m = 1000$ and $m = 5000$, and show that the empirical RMSEs of this estimator reflect good agreement with the theoretical standard deviations with convergence rate $m^{1/2}$ for $m = 5000$.

Chapter 5

A generalized nonlinear model for long memory conditional heteroscedasticity

In this chapter, we study the existence and properties of a stationary solution of the ARCH-type equation $r_t = \zeta_t \sigma_t$, where ζ_t are standardized i.i.d. random variables and the conditional variance satisfies an AR(1) equation $\sigma_t^2 = Q^2\big(a + \sum_{j=1}^{\infty} b_j r_{t-j}\big) + \gamma \sigma_{t-1}^2$ with a Lipschitz function $Q(x)$ and real parameters a, γ, b_j. We extend the model and the results by Doukhan, Grublytė and Surgailis [22] from the case $\gamma = 0$ to the case $0 < \gamma < 1$. We also obtain a new condition for the existence of higher moments of r_t, which does not include the Rosenthal constant. In particular, when Q is the square root of a quadratic polynomial, we prove that r_t can exhibit a leverage effect and long memory. The parametric QML estimation for the latter model (called the Generalized Quadratic ARCH model, GQARCH) is considered in Chapter 4 of this dissertation.

5.1 Introduction

Doukhan *et al.* [22] discussed the existence of a stationary solution of the conditionally heteroscedastic equation

$$r_t \;=\; \zeta_t\sigma_t, \qquad \sigma_t^2 \;=\; Q^2\left(a + \sum_{j=1}^{\infty} b_j r_{t-j}\right), \tag{5.1}$$

where $\{\zeta_t, t \in \mathbb{Z}\}$ are standardized i.i.d. random variables, a, b_j, are real parameters and $Q(x)$ is a Lipschitz function of real variable $x \in \mathbb{R}$. The most important case of (5.1) probably is

$$Q(x) = \sqrt{c^2 + x^2}, \tag{5.2}$$

where $c \geq 0$ is a parameter. Models (5.1)–(5.2) include the classical Asymmetric ARCH(1) of Engle [25] and the Linear ARCH (LARCH) model of Robinson [62]:

$$r_t \;=\; \zeta_t\sigma_t, \qquad \sigma_t \;=\; a + \sum_{j=1}^{\infty} b_j r_{t-j}. \tag{5.3}$$

Giraitis *et al.* [32] proved that the squared stationary solution $\{r_t^2, t \in \mathbb{Z}\}$ of the LARCH model in Equation (5.3) with b_j decaying as $j^{d-1}, 0 < d < 1/2$, may have long memory autocorrelations. The leverage effect in the LARCH model was discussed in detail by Giraitis *et al.* [33]. Doukhan *et* *al.* [22] extended the above properties of the LARCH model (long memory and leverage) to the model of (5.1)–(5.2) with $c > 0$, or strictly positive volatility.

In this chapter, we extend the results of Doukhan *et al.* [22] to a more general class of volatility forms:

$$r_t = \zeta_t \sigma_t, \qquad \sigma_t^2 = Q^2\left(a + \sum_{j=1}^{\infty} b_j r_{t-j}\right) + \gamma \sigma_{t-1}^2, \tag{5.4}$$

where $\{\zeta_t, t \in \mathbb{Z}\}, a, b_j, Q(x)$, are as in (5.1) and $0 < \gamma < 1$ is a parameter. The inclusion of lagged σ_{t-1}^2 in (5.4) helps to reduce very sharp peaks and clustering of volatility which occur in trajectory of models (5.1)–(5.2) near the threshold $c > 0$. The generalization from (5.1) to (5.4) is similar to that from ARCH to GARCH models (see, e.g., Engle [24] and Bollerslev [10]), particularly, equation (5.4) with $Q(x)$ in (5.2) and $b_j = 0, j \geq 2$, reduces to the Asymmetric GARCH(1,1) of Engle [25]:

$$\sigma_t^2 = c^2 + (a + br_{t-1})^2 + \gamma \sigma_{t-1}^2.$$

Let us describe the main results of this chapter. Section 5.2 (Theorems 5.1 and 5.2) obtains sufficient conditions for the existence of a stationary solution of (5.4) with $\mathrm{E}|r_t|^p < \infty$ and $\gamma \in [0, 1)$. Theorem 5.1 extends the corresponding result by Doukhan *et al.* [22] (Theorem 4) from $\gamma = 0$ to $\gamma > 0$. Theorem 5.2 is new even in the case $\gamma = 0$, by providing an explicit sufficient condition (5.24) for higher-order even moments ($p = 4, 6, \ldots$), which does not involve the absolute constant in Burkholder-Rosenthal inequality (5.11). Condition (5.24) coincides with the corresponding moment condition for the LARCH model and is important for statistical applications, see

Remark 5.2. Sections 5.3–5.4 deal exclusively with the case of quadratic Q^2 in (5.2), referred to as the Generalized Quadratic ARCH (GQARCH) model. Theorem 5.3 (Section 5.3) obtains long memory properties of the squared process $\{r_t^2, t \in \mathbb{Z}\}$ of the GQARCH model with $\gamma \in (0,1)$ and coefficients b_j decaying regularly as $b_j \sim \beta j^{d-1}$, $j \to \infty$, $0 < d < 1/2$. Similar properties were established by Doukhan *et al.* [22] for the GQARCH model with $\gamma = 0$ and for the LARCH model (5.3) by Giraitis, Robinson and Surgailis [32], Giraitis, Leipus, Robinson and Surgailis [33]. The quasi-maximum likelihood estimation for the parametric GQARCH model with long memory was studied in Chapter 4 of this dissertation (see also Grublytė, Surgailis and Škarnulis [41]). See the review paper by Giraitis, Leipus and Surgailis [35] and Chapter 3 of this dissertation (also, Giraitis, Surgailis and Škarnulis [37]) for issues related with long memory ARCH modeling. Section 5.4 extends to the GQARCH model the leverage effect discussed by Doukhan *et al.* [22] and Giraitis *et al.* [33].

5.2 Stationary solution

Denote $|\mu|_p := \mathrm{E}|\zeta_0|^p \ (p > 0)$, $\mu_p := \mathrm{E}\zeta_0^p \ (p = 1, 2, \dots)$ and let

$$X_t := \sum_{s<t} b_{t-s} r_s. \tag{5.5}$$

Since $0 \le \gamma < 1$, equations in (5.4) yield

$$\sigma_t^2 = \sum_{\ell=0}^{\infty} \gamma^\ell Q^2(a + X_{t-\ell}) \quad \text{and} \quad r_t = \zeta_t \sqrt{\sum_{\ell=0}^{\infty} \gamma^\ell Q^2(a + X_{t-\ell})}. \tag{5.6}$$

In other words, stationary solution of (5.4), or

$$r_t \;=\; \zeta_t \sqrt{\sum_{\ell=0}^{\infty} \gamma^\ell Q^2 \left(a + \sum_{j=1}^{\infty} b_j r_{t-\ell-j} \right)}, \tag{5.7}$$

can be defined via (5.5), or stationary solution of

$$X_t \;:=\; \sum_{s<t} b_{t-s} \zeta_s \sqrt{\sum_{\ell=0}^{\infty} \gamma^\ell Q^2 (a + X_{s-\ell})}, \tag{5.8}$$

and vice versa.

In Theorem 5.1, we assume that Q in (5.6) is a Lipschitz function, that is, there exists a constant $\mathrm{Lip}_Q > 0$, such that

$$|Q(x) - Q(y)| \le \mathrm{Lip}_Q |x - y|, \qquad x, y \in \mathbb{R}. \tag{5.9}$$

Note that inequality (5.9) implies the bound

$$Q^2(x) \le c_1^2 + c_2^2 x^2, \qquad x \in \mathbb{R}, \tag{5.10}$$

where $c_1 \ge 0,\ c_2 \ge \mathrm{Lip}_Q$ and c_2 can be chosen arbitrarily close to Lip_Q.

Let us give some formal definitions. As in Chapter 4, let $\mathcal{F}_t = \sigma(\zeta_s, s \le t), t \in \mathbb{Z}$, be the sigma-field generated by $\zeta_s, s \le t$. A random process $\{u_t, t \in \mathbb{Z}\}$ is called *adapted* (respectively, *predictable*) if u_t is $\mathcal{F}_t$-measurable for each $t \in \mathbb{Z}$ (respectively, u_t is $\mathcal{F}_{t-1}$-measurable for each $t \in \mathbb{Z}$).

Definition 5.1. *Let $p > 0$ be an arbitrary real number.*

(i) *By L^p-solution of (5.6) or/and (5.7) we mean an adapted process $\{r_t, t \in \mathbb{Z}\}$ with $\mathrm{E}|r_t|^p < \infty$ such that, for any $t \in \mathbb{Z}$, the series $X_t = \sum_{j=1}^{\infty} b_j r_{t-j}$ converges* *in* L^p*, the series* $\sigma_t^2 = \sum_{\ell=0}^{\infty} \gamma^\ell Q^2(a + X_{t-\ell})$ *converges in* $L^{p/2}$ *and (5.7) holds.*

(ii) *By* L^p*-solution of (5.8) we mean a predictable process* $\{X_t, t \in \mathbb{Z}\}$ *with* $\mathrm{E}|X_t|^p < \infty$ *such that, for any* $t \in \mathbb{Z}$*, the series* $\sigma_t^2 = \sum_{\ell=0}^{\infty} \gamma^\ell Q^2(a + X_{t-\ell})$ *converges in* $L^{p/2}$*, the series* $\sum_{s<t} b_{t-s}\zeta_s\sigma_s$ *converges in* L^p *and (5.8) holds.*

Define

$$B_p := \begin{cases} \sum_{j=1}^{\infty} |b_j|^p, & 0 < p < 2, \\ \big(\sum_{j=1}^{\infty} b_j^2\big)^{p/2}, & p \geq 2, \end{cases} \qquad B_{p,\gamma} := \begin{cases} B_p/(1-\gamma^{p/2}), & 0 < p < 2, \\ B_p/(1-\gamma)^{p/2}, & p \geq 2. \end{cases}$$

Note $B_p = B_{p,0}$. Similarly to Doukhan *et al.* [22], we use the following moment inequality, see Burkholder [14], Bahr and Esséen [69], Rosenthal [64].

Proposition 5.1. *Let* $\{Y_j, j \geq 1\}$ *be a sequence of random variables such that* $\mathrm{E}|Y_j|^p < \infty$ *for some* $p > 0$*. If* $p > 1$*, we additionally assume that* $\{Y_j\}$ *is a martingale difference sequence:* $\mathrm{E}[Y_j|Y_1, \ldots, Y_{j-1}] = 0$*,* $j = 2, 3, \ldots$*. Then there exists a constant* $K_p \geq 1$ *depending only on* p *and such that*

$$\mathrm{E}\Big|\sum_{j=1}^{\infty} Y_j\Big|^p \;\leq\; K_p \begin{cases} \sum_{j=1}^{\infty} \mathrm{E}|Y_j|^p, & 0 < p \leq 2, \\ \big(\sum_{j=1}^{\infty} (\mathrm{E}|Y_j|^p)^{2/p}\big)^{p/2}, & p > 2. \end{cases} \tag{5.11}$$

Proposition 5.2 states that equations (5.7) and (5.8) are equivalent in the sense that by solving one of these equations, one readily obtains a solution to the other one.

Proposition 5.2. *Let* Q *be a measurable function satisfying (5.10) with some* $c_1, c_2 \geq 0$*, and* $\{\zeta_t, t \in \mathbb{Z}\}$ *be an i.i.d. sequence with* $|\mu|_p = \mathrm{E}|\zeta_0|^p < \infty$ *and satisfying* $\mathrm{E}\zeta_0 = 0$ *for* $p > 1$*. In addition, assume* $B_p < \infty$ *and* $0 \leq \gamma < 1$*.*

(i) *Let* $\{X_t, t \in \mathbb{Z}\}$ *be a stationary* L^p*-solution of (5.8) and let*

$$\sigma_t := \sqrt{\sum_{\ell=0}^{\infty} \gamma^\ell Q^2(a + X_{t-\ell})}.$$

Then $\{r_t = \zeta_t \sigma_t\}$ *in (5.6) is a stationary* L^p*-solution of equation (5.7) and*

$$\mathrm{E}|r_t|^p \ \le \ C(1 + \mathrm{E}|X_t|^p). \tag{5.12}$$

Moreover, for $p > 1$, $\{r_t, \mathcal{F}_t, t \in \mathbb{Z}\}$ *is a martingale difference sequence with*

$$\mathrm{E}[r_t|\mathcal{F}_{t-1}] = 0, \qquad \mathrm{E}[|r_t|^p|\mathcal{F}_{t-1}] = |\mu|_p \sigma_t^p. \tag{5.13}$$

(ii) *Let* $\{r_t, t \in \mathbb{Z}\}$ *be a stationary* L^p*-solution of (5.7). Then* $\{X_t, t \in \mathbb{Z}\}$*, defined in (5.5), is a stationary* L^p*-solution of equation (5.8) such that*

$$\mathrm{E}|X_t|^p \ \le \ C\mathrm{E}|r_t|^p.$$

Moreover, for $p \ge 2$*,*

$$\mathrm{E}[X_t X_0] \ = \ \mathrm{E}r_0^2 \sum_{s=1}^{\infty} b_{t+s} b_s, \qquad t = 0, 1, \dots.$$

Proof. (i) First, let $0 < p \le 2$. Then

$$\mathrm{E}|\sigma_t|^p = \mathrm{E}|\sigma_t^2|^{p/2} \le \sum_{\ell=0}^{\infty} |\gamma^{p/2}|^\ell \mathrm{E}|Q(a + X_{t-\ell})|^p < \infty.$$

Hence, using inequality (5.10), the fact that $\{X_t, t \in \mathbb{Z}\}$ is predictable and

$$|Q(a+X_{t-\ell})|^p \le |c_1^2 + c_2^2(a+X_{t-\ell})|^{p/2} \le C(1+|a+X_{t-\ell}|^p) \le C(1+|X_{t-\ell}|^p),$$

we obtain

$$\begin{aligned}\mathrm{E}|r_t|^p &= |\mu|_p \mathrm{E}|\sigma_t|^p \le C\sum_{\ell=0}^{\infty}|\gamma^{p/2}|^{\ell}(1+\mathrm{E}|X_{t-\ell}|^p)\\ &\le C(1+\mathrm{E}|X_t|^p) < \infty,\end{aligned}$$

proving (5.12) for $p \le 2$. Next, let $p > 2$. Then

$$\mathrm{E}|\sigma_t|^p \le \left(\sum_{\ell=0}^{\infty}\gamma^{\ell}\mathrm{E}^{2/p}|Q(a+X_t)|^p\right)^{p/2} \le C\mathrm{E}|Q(a+X_t)|^p,$$

by stationarity and Minkowski's inequality and hence (5.12) follows using the same argument as above. Clearly, for $p > 1$, $\{r_t = \zeta_t\sigma_t\}$ is a martingale difference sequence and satisfies equations in (5.13). Then, the convergence in L^p of the series in equation (5.5) follows from (5.12) and Proposition 5.1:

$$\mathrm{E}\Big|\sum_{j=1}^{\infty} b_j r_{t-j}\Big|^p \le C\left\{\begin{array}{ll}\sum_{j=1}^{\infty}|b_j|^p, & 0<p\le 2\\ \left(\sum_{j=1}^{\infty}b_j^2\right)^{p/2}, & p>2\end{array}\right\} = CB_p < \infty.$$

In particular,

$$\zeta_t\sqrt{\sum_{\ell=0}^{\infty}\gamma^{\ell}Q^2(a+\sum_{s<t}b_{t-\ell-s}r_s)} = \zeta_t\sqrt{\sum_{\ell=0}^{\infty}\gamma^{\ell}Q^2(a+X_{t-\ell})} = r_t,$$

by the definition of r_t. Hence, $\{r_t, t\in\mathbb{Z}\}$ is a L^p-solution of equation (5.7). Stationarity of $\{r_t, t\in\mathbb{Z}\}$ follows from stationarity of $\{X_t, t\in\mathbb{Z}\}$.

(ii) Since $\{r_t, t \in \mathbb{Z}\}$ is a L^p-solution of equation (5.7), so

$$r_t = \zeta_t \sigma_t = \zeta_t \sqrt{\sum_{\ell=0}^{\infty} \gamma^{\ell} Q^2(a + X_{t-\ell})},$$

with X_t defined in (5.5) and $\{X_t, t \in \mathbb{Z}\}$ satisfies equation (5.5), where the series converges in L^p. The rest follows as in Doukhan *et al.* [22], proof of Proposition 2. □

Remark 5.1. Let $p \geq 2$ and $|\mu|_p < \infty$, then by inequality (5.11), $\{r_t, t \in \mathbb{Z}\}$ being a stationary L^p-solution of (5.6) is equivalent to $\{r_t, t \in \mathbb{Z}\}$ being a stationary L^2-solution of (5.6) with $\mathrm{E}|r_0|^p < \infty$. Similarly, if Q and $\{\zeta_t, t \in \mathbb{Z}\}$ satisfy the conditions of Proposition 5.2 and $p \geq 2$, then $\{X_t, t \in \mathbb{Z}\}$, being a stationary L^p-solution of (5.5), is equivalent to $\{X_t, t \in \mathbb{Z}\}$ being a stationary L^2-solution of (5.5) with $\mathrm{E}|X_0|^p < \infty$. See also Doukhan *et al.* [22], Remark 1.

Theorem 5.1. *Let $\{\zeta_t, t \in \mathbb{Z}\}$ satisfy the conditions of Proposition 5.2 and Q satisfy the Lipschitz condition in (5.9).*

(i) *Let $p > 0$ and*

$$K_p^{1/p} \, |\mu|_p^{1/p} \, \mathrm{Lip}_Q \, B_{p,\gamma}^{1/p} < 1, \tag{5.14}$$

where K_p is the absolute constant from the moment inequality in (5.11). Then there exists a unique stationary L^p-solution $\{X_t, t \in \mathbb{Z}\}$ of (5.8) and

$$\mathrm{E}|X_t|^p \ \leq \frac{C(p,Q)|\mu|_p B_p}{1 - K_p|\mu|_p \mathrm{Lip}_Q^p B_{p,\gamma}}, \tag{5.15}$$

where $C(p,Q) < \infty$ depends only on p and c_1, c_2 in inequality (5.10).

(ii) *Assume, in addition, that $Q^2(x) = c_1^2 + c_2^2 x^2$, where $c_i \geq 0$, $i = 1, 2$, and*

$\mu_2 = \mathrm{E}\zeta_0^2 = 1$. *Then* $c_2^2 B_{2,\gamma} < 1$ *is a necessary and sufficient condition for the existence of a stationary* L^2*-solution* $\{X_t, t \in \mathbb{Z}\}$ *of equation (5.8) with* $a \neq 0$.

Proof. (i) We follow the proof of Theorem 4 by Doukhan *et al.* [22]. For $n \in \mathbb{N}$ we recurrently define a solution of equation (5.8) with zero initial condition at $t \leq -n$ as

$$X_t^{(n)} \;:=\; \begin{cases} 0, & t \leq -n, \\ \sum_{s=-n}^{t-1} b_{t-s}\zeta_s\sigma_s^{(n)}, & t > -n, \quad t \in \mathbb{Z}, \end{cases} \tag{5.16}$$

where $\sigma_s^{(n)} := \sqrt{\sum_{\ell=0}^{n+s} \gamma^\ell Q^2(a + X_{s-\ell}^{(n)})}$. Let us show that $\{X_t^{(n)}, t \in \mathbb{Z}\}$ converges in L^p to a stationary L^p-solution $\{X_t, t \in \mathbb{Z}\}$ as $n \to \infty$.

First, let $0 < p \leq 2$. Let $m > n \geq 0$. Then, by inequality (5.11), for any $t > -m$ we have that

$$\begin{aligned} \mathrm{E}|X_t^{(m)} - X_t^{(n)}|^p \;&\leq\; K_p|\mu|_p\Big\{\sum_{-m\leq s<-n} |b_{t-s}|^p \mathrm{E}|\sigma_s^{(m)}|^p + \\ &\quad + \sum_{-n\leq s<t} |b_{t-s}|^p \mathrm{E}|\sigma_s^{(m)} - \sigma_s^{(n)}|^p\Big\} \\ &=: \; K_p|\mu|_p\big\{S'_{m,n} + S''_{m,n}\big\}. \end{aligned} \tag{5.17}$$

Using $|Q(a+x)|^p \leq C + c_3^p|x|^p, x \in \mathbb{R}$, with $c_3 > c_2 > \mathrm{Lip}_Q$ arbitrarily close to Lip_Q, see Doukhan *et al.* [22], proof of Theorem 4, we obtain

$$S'_{m,n} \;\leq\; \sum_{-m\leq s<-n} |b_{t-s}|^p \sum_{\ell=0}^{m+s} \gamma^{p\ell/2}\left(C + c_3^p \mathrm{E}|X_{s-\ell}^{(m)}|^p\right). \tag{5.18}$$

Next, using $|(\sum_{i>0}\gamma^i x_i^2)^{1/2} - (\sum_{i>0}\gamma^i y_i^2)^{1/2}| \leq (\sum_{i>0}\gamma^i (x_i - y_i)^2)^{1/2}$ we obtain

$$
\begin{aligned}
|\sigma_s^{(m)} - \sigma_s^{(n)}| \ \le \ & \left(\sum_{\ell=0}^{s+n} \gamma^\ell \big(Q(a + X_{s-\ell}^{(m)}) - Q(a + X_{s-\ell}^{(n)})\big)^2 + \right. \\
& \left. + \sum_{\ell=s+n+1}^{s+m} \gamma^\ell Q^2(a + X_{s-\ell}^{(m)}) \right)^{1/2} . \qquad (5.19)
\end{aligned}
$$

Hence from the Lipschitz condition in (5.9) we have that

$$
\begin{aligned}
S''_{m,n} \ \le \ & \sum_{-n \le s < t} |b_{t-s}|^p \left(\sum_{\ell=0}^{s+n} \gamma^{p\ell/2} \mathrm{Lip}_Q^p \mathrm{E}|X_{s-\ell}^{(m)} - X_{s-\ell}^{(n)}|^p + \right. \\
& \left. + \sum_{\ell=s+n+1}^{s+m} \gamma^{p\ell/2} (C + c_3^p \mathrm{E}|X_{s-\ell}^{(m)})|^p) \right) .
\end{aligned}
$$

Combining (5.17) and the above bounds, we obtain

$$
\begin{aligned}
\mathrm{E}|X_t^{(m)} - X_t^{(n)}|^p \ \le \ & K_p|\mu|_p \left(c_3^p \sum_{-m \le s < t} |b_{t-s}|^p \sum_{\ell=0}^{s+m} \gamma^{p\ell/2} \mathrm{E}|X_{s-\ell}^{(m)} - X_{s-\ell}^{(n)}|^p \right. \\
& + \ C \sum_{-m \le s < -n} |b_{t-s}|^p \sum_{\ell=0}^{s+m} \gamma^{p\ell/2} + \\
& \left. + \ C \sum_{-n \le s < t} |b_{t-s}|^p \sum_{\ell=s+n+1}^{s+m} \gamma^{p\ell/2} \right) \\
\le \ & C K_p |\mu|_p \kappa_{t+n,\gamma}^p + \\
& + \ K_p |\mu|_p c_3^p \sum_{-m \le s < t} b_{t-s,\gamma}^p \mathrm{E}|X_s^{(n)} - X_s^{(m)}|^p, \qquad (5.20)
\end{aligned}
$$

where $b_{s,\gamma}^p := \sum_{j=0}^{s-1} \gamma^{jp/2} |b_{s-j}|^p$, $s \ge 0$, and

$$
\kappa_{t+n} := C(1-\gamma^{p/2})^{-1} K_p |\mu|_p \left(\sum_{j > t+n} |b_j|^p + b_{t+n,\gamma}^p \right) \to 0 \, (n \to \infty).
$$

Iterating inequality (5.20) as in Doukhan *et al.* [22], (6.3), and using

$$K_p|\mu|_p c_3^p \sum_{s<t} b_{t-s,\gamma}^p = K_p|\mu|_p c_3^p B_{p,\gamma} < 1,$$

we obtain $\lim_{m,n\to\infty} \mathrm{E}|X_t^{(m)} - X_t^{(n)}|^p = 0$, and hence the existence of X_t such that $\lim_{n\to\infty} \mathrm{E}|X_t^{(n)} - X_t|^p = 0$ and satisfying the bound in (5.15).

Next, consider the case $p > 2$. Let $m > n \geq 0$. Then by inequality (5.11) for any $t > -m$, we have that

$$\begin{aligned}\mathrm{E}|X_t^{(m)} - X_t^{(n)}|^p &\leq K_p|\mu|_p \Bigg(\sum_{-m\leq s<-n} b_{t-s}^2 \mathrm{E}^{2/p}|\sigma_s^{(m)}|^p + \\ &\quad + \sum_{-n\leq s<t} b_{t-s}^2 \mathrm{E}^{2/p}|\sigma_s^{(m)} - \sigma_s^{(n)}|^p \Bigg)^{p/2} \\ &=: K_p|\mu|_p \big(R'_{m,n} + R''_{m,n}\big)^{p/2}. \end{aligned} \tag{5.21}$$

Similar to (5.18),

$$\begin{aligned} R'_{m,n} &\leq \sum_{s=-m}^{-n-1} b_{t-s}^2 \sum_{\ell=0}^{m+s} \gamma^\ell \mathrm{E}^{2/p}|Q(a + X_{s-\ell}^{(m)})|^p \\ &\leq \sum_{s=-m}^{-n-1} b_{t-s}^2 \sum_{\ell=0}^{m+s} \gamma^\ell (C + c_3^2 \mathrm{E}^{2/p}|X_{s-\ell}^{(m)}|^p), \end{aligned}$$

and using inequality (5.19),

$$\begin{aligned} R''_{m,n} &\leq \sum_{-n\leq s<t} b_{t-s}^2 \mathrm{E}^{2/p} \Bigg| \sum_{\ell=0}^{s+n} \gamma^\ell (Q(a + X_{s-\ell}^{(m)}) - Q(a + X_{s-\ell}^{(n)}))^2 + \\ &\quad + \sum_{\ell=s+n+1}^{s+m} \gamma^\ell Q^2(a + X_{s-\ell}^{(m)}) \Bigg|^{p/2} \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{-n\leq s<t} b_{t-s}^2 \left(\sum_{\ell=0}^{s+n} \gamma^\ell \mathrm{E}^{2/p} |Q(a+X_{s-\ell}^{(m)}) - Q(a+X_{s-\ell}^{(n)})|^p + \right. \\
&\quad \left. + \sum_{\ell=s+n+1}^{s+m} \gamma^\ell \mathrm{E}^{2/p} |Q(a+X_{s-\ell}^{(m)})|^p \right) \\
&\leq \sum_{-n\leq s<t} b_{t-s}^2 \left(\mathrm{Lip}_Q^2 \sum_{\ell=0}^{s+n} \gamma^\ell \mathrm{E}^{2/p} |X_{s-\ell}^{(m)} - X_{s-\ell}^{(n)}|^p + \right. \\
&\quad \left. + \sum_{\ell=s+n+1}^{s+m} \gamma^\ell (C + c_3^2 \mathrm{E}^{2/p} |X_{s-\ell}^{(m)}|^p) \right).
\end{aligned}
$$

Consequently,

$$
\mathrm{E}^{2/p}|X_t^{(m)} - X_t^{(n)}|^p \leq \kappa_{t+n} + K_p^{2/p}|\mu|_p^{2/p} c_3^2 \sum_{-m\leq s<t} b_{t-s,\gamma}^2 \mathrm{E}^{2/p}|X_s^{(m)}) - X_s^{(n)}|^p,
$$

where $\kappa_{t+n} := C(1-\gamma)^{-1} K_p^{2/p} |\mu|_p^{2/p} (\sum_{j>t+n} b_j^2 + b_{t+n,\gamma}^2) \to 0\, (n\to\infty)$. By iterating the last displayed equation and using

$$
K_p^{2/p}|\mu|_p^{2/p} c_3^2 \sum_{j=1} b_{j,\gamma}^2 = K_p^{2/p}|\mu|_p^{2/p} c_3^2 B_2/(1-\gamma) < 1,
$$

we obtain $\lim_{m,n\to\infty} \mathrm{E}^{2/p}|X_t^{(m)} - X_t^{(n)}|^p = 0$ and hence the existence of X_t such that $\lim_{n\to\infty} \mathrm{E}|X_t^{(n)} - X_t|^p = 0$ and satisfying the bound in (5.15). The rest of the proof of part (i) is similar as in Doukhan *et al.* [22], proof of Theorem 1, and we omit the details.

(ii) Note that $Q(x) = \sqrt{c_1^2 + c_2^2 x^2}$ is a Lipschitz function and satisfies (5.9) with $\mathrm{Lip}_Q = c_2$. Hence by $K_2 = 1$ and part (i), a unique L^2-solution $\{X_t, t\in\mathbb{Z}\}$ of equation (5.8) under the condition $c_2^2 B_{2,\gamma} = c_2^2 B_2/(1-\gamma) < 1$ exists. To show the necessity of the last condition, let $\{X_t, t\in\mathbb{Z}\}$ be a stationary L^2-solution of equation (5.8). Then

$$\begin{aligned}
\mathrm{E}X_t^2 &= \sum_{s<t} b_{t-s}^2 \sum_{\ell=0}^{\infty} \gamma^\ell \mathrm{E}Q^2(a+X_{s-\ell}) \\
&= \sum_{s<t} b_{t-s}^2 \sum_{\ell=0}^{\infty} \gamma^\ell \mathrm{E}\big(c_1^2 + c_2^2(a+X_{s-\ell}^2\big) \\
&= (B_2/(1-\gamma))\big(c_1^2 + c_2^2(a^2+\mathrm{E}X_t^2)\big) \; > c_2^2(B_2/(1-\gamma))\mathrm{E}X_t^2,
\end{aligned}$$

since $a \neq 0$. Hence, $c_2^2 B_2/(1-\gamma) < 1$ unless $\mathrm{E}X_t^2 = 0$, or $\{X_t = 0\}$ is a trivial process. Clearly, equation (5.8) admits a trivial solution if and only if $0 = Q(a) = \sqrt{c_1^2 + c_2^2 a^2} = 0$, or $c_1 = c_2 = 0$. This proves part (ii) and the theorem. □

Remark 5.2. Theorem 5.1 extends Theorem 4 by Doukhan *et al.* [22] from $\gamma = 0$ to $\gamma > 0$. A major shortcoming of Theorem 5.1 and the above mentioned result by Doukhan *et al.* [22] is the presence of universal constant K_p in condition 5.14. The upper bound of K_p given by Osękowski [23] leads to restrictive conditions on $B_{p,\gamma}$ in inequality (5.14) for the existence of the L^p-solution, $p > 2$. For example, for $p = 4$, the above mentioned bound of Osękowski [23] gives

$$K_4\mu_4 B_2^2/(1-\gamma)^2 \le (27.083)^4 \mu_4 B_2^2/(1-\gamma)^2 < 1, \tag{5.22}$$

requiring $B_2 = \sum_{j=1}^{\infty} b_j^2$ to be very small. Since statistical inference based on "observable" squares $r_t^2, 1 \le t \le n,$ usually requires the existence of $\mathrm{E}r_t^4$ and higher moments of r_t (see, e.g., Grublytė *et al.* [41]), there is a necessity to derive less restrictive conditions for the existence of these moments which do not involve the Rosenthal constant K_p. This is achieved in Theorem 5.2. Particularly, for $\gamma = 0, \mathrm{Lip}_Q = 1,$ sufficient condition (5.24) of Theorem 5.2 for the existence of $\mathrm{E}r_t^p, p \geq 2$, even becomes

$$\sum_{j=2}^{p} \binom{p}{j} |\mu_j| \sum_{k=1}^{\infty} |b_k|^j \ < \ 1. \tag{5.23}$$

Condition (5.23) coincides with the corresponding condition in the LARCH case in Giraitis *et al.* [33], Proposition 3. Moreover, conditions (5.23) and (5.24) apply to more general classes of ARCH models in (5.1) and (5.4), to which the specific Volterra series techniques used by Giraitis, Robinson and Surgailis [32], and Giraitis, Leipus, Robinson and Surgailis [33] are not applicable. In the particular case $p = 4$, condition (5.23) becomes

$$6B_2 + 4|\mu_3| \sum_{k=1}^{\infty} |b_k|^3 + \mu_4 \sum_{k=1}^{\infty} |b_k|^4 < 1,$$

which seems to be much better than condition (5.22) based on Theorem 5.1.

Theorem 5.2. *Let $\{\zeta_t, t \in \mathbb{Z}\}$ satisfy the conditions of Proposition 5.2 and Q satisfy the Lipschitz condition (5.9). Let $p = 2, 4, \ldots,$ be even and*

$$\sum_{j=2}^{p} \binom{p}{j} |\mu_j| \mathrm{Lip}_Q^j \sum_{k=1}^{\infty} |b_k|^j \ < \ (1-\gamma)^{p/2}. \tag{5.24}$$

Then there exists a unique stationary L^p-solution $\{X_t, t \in \mathbb{Z}\}$ of equation (5.8).

Proof. For $p = 2$, condition (5.24) agrees with $\mathrm{Lip}_Q^2 B_{2,\gamma} < 1$ or condition (5.14), so we assume $p \geq 4$ in the subsequent proof. In the latter case, inequality (5.24) implies $\mathrm{Lip}_Q^2 B_{2,\gamma} < 1$ and the existence of a stationary L^2-solution $\{X_t, t \in \mathbb{Z}\}$ of equation (5.8). It suffices to show that the above L^2-solution satisfies $\mathrm{E}X_t^p < \infty$.

To this end, similarly as in the proof of Theorem 5.1 (i), consider the solution $\{X_t^{(n)}\}$ with zero initial condition at $t \le -n$ as defined in (5.16). Let $\sigma_t^{(n)} := 0, t < -n$. Since $\mathrm{E}(X_t^{(n)} - X_t)^2 \to 0$ $(n \to \infty)$, by Fatou's lemma it suffices to show that under condition (5.24)

$$\mathrm{E}(X_t^{(n)})^p \ < \ C, \tag{5.25}$$

where the constant $C < \infty$ does not depend on t, n.

Since p is even, for any $t > -n$ we have that

$$\begin{aligned}\mathrm{E}(X_t^{(n)})^p &= \sum_{s_1,\ldots,s_p=-n}^{t-1} \mathrm{E}\big[b_{t-s_1}\zeta_{s_1}\sigma_{s_1}^{(n)} \cdots b_{t-s_p}\zeta_{s_p}\sigma_{s_p}^{(n)}\big] \\ &= \sum_{j=2}^{p} \binom{p}{j} \sum_{s=-n}^{t-1} b_{t-s}^j \mu_j \mathrm{E}\Big[(\sigma_s^{(n)})^j \Big(\sum_{u=-n}^{s-1} b_{t-u}\zeta_u\sigma_u^{(n)}\Big)^{p-j}\Big].\end{aligned} \tag{5.26}$$

Hence using Hölder's inequality:

$$\begin{aligned}|\mathrm{E}\xi^j\eta^{p-j}| &\le c^j \mathrm{E}^{j/p}|\xi/c|^p \mathrm{E}^{(p-j)/p}|\eta|^p \\ &\le c^j \left[\frac{j}{pc^p}\mathrm{E}|\xi|^p + \frac{p-j}{p}\mathrm{E}|\eta|^p\right], \quad 1 \le j \le p,\ c > 0,\end{aligned}$$

we obtain

$$\begin{aligned}\mathrm{E}(X_t^{(n)})^p &\le \sum_{j=2}^{p} \binom{p}{j} |\mu_j| c_3^j \sum_{s=-n}^{t-1} |b_{t-s}^j| \left\{ \frac{j}{pc_3^p}\mathrm{E}(\sigma_s^{(n)})^p + \right. \\ &\qquad \left. + \frac{p-j}{p}\mathrm{E}\left(\sum_{u=-n}^{s-1} b_{t-u}\zeta_u\sigma_u^{(n)}\right)^p \right\} \\ &= \sum_{s=-n}^{t-1} \beta_{1,t-s}\mathrm{E}(\sigma_s^{(n)}/c_3)^p + \sum_{s=-n}^{t-1} \beta_{2,t-s}\mathrm{E}\big(X_{t,s}^{(n)}\big)^p,\end{aligned} \tag{5.27}$$

where $X^{(n)}_{t,s} := \sum_{u=-n}^{s-1} b_{t-u}\zeta_u\sigma^{(n)}_u$, $c_3 > \mathrm{Lip}_Q$, and where

$$\beta_{1,t-s} \;:=\; \sum_{j=2}^{p} \frac{j}{p}\binom{p}{j}|b^j_{t-s}||\mu_j|c^j_3, \quad \beta_{2,t-s} \;:=\; \sum_{j=2}^{p} \frac{p-j}{p}\binom{p}{j}|b^j_{t-s}||\mu_j|c^j_3.$$

The last expectation in (5.27) can be evaluated similarly to (5.26)–(5.27):

$$\begin{aligned} \mathrm{E}\big(X^{(n)}_{t,s}\big)^p &= \sum_{j=2}^{p}\binom{p}{j}\sum_{u=-n}^{s-1} b^j_{t-u}\mu_j \mathrm{E}\left[(\sigma^{(n)}_u)^j\left(\sum_{v=-n}^{u-1} b_{t-v}\zeta_v\sigma^{(n)}_v\right)^{p-j}\right] \\ &\le \sum_{u=-n}^{s-1}\beta_{1,t-u}\mathrm{E}(\sigma^{(n)}_u/c_3)^p + \sum_{u=-n}^{s-1}\beta_{2,t-u}\mathrm{E}\big(X^{(n)}_{t,u}\big)^p. \end{aligned}$$

Proceeding recurrently with the above evaluation results in the inequality:

$$\mathrm{E}(X^{(n)}_t)^p \;\le\; \sum_{s=-n}^{t-1}\widetilde{\beta}_{t-s}\mathrm{E}(\sigma^{(n)}_s/c_3)^p, \tag{5.28}$$

where

$$\widetilde{\beta}_{t-s} \;:=\; \beta_{1,t-s}\left(1 + \sum_{k=1}^{t-s-1}\sum_{s<u_k<\cdots<u_1<t}\beta_{2,t-u_1}\cdots\beta_{2,t-u_k}\right).$$

Let $\beta_i := \sum_{t=1}^{\infty}\beta_{i,t}$, $i = 1, 2$, $\widetilde{\beta} := \sum_{t=1}^{\infty}\widetilde{\beta}_t$. By assumption (5.24),

$$\beta_1 + \beta_2 \;=\; \sum_{j=2}^{p}\binom{p}{j}|\mu_j|c^j_3\sum_{k=1}^{\infty}|b_k|^j \;<\; (1-\gamma)^{p/2},$$

whenever $\sigma_3 - \mathrm{Lip}_Q > 0$ is small enough, and, therefore,

$$\begin{aligned} \frac{\widetilde{\beta}}{(1-\gamma)^{p/2}} &\le \frac{1}{(1-\gamma)^{p/2}}\sum_{t=1}^{\infty}\beta_{1,t}\left(1+\sum_{k=1}^{\infty}\beta^k_2\right) \\ &= \frac{1}{(1-\gamma)^{p/2}}\frac{\beta_1}{1-\beta_2} \;<\; 1. \end{aligned} \tag{5.29}$$

Next, let us estimate the expectation on the r.h.s. of inequality (5.28) in terms of expectations on the l.h.s. Using inequality (5.10) and Minkowski's inequalities, we obtain

$$\begin{aligned} \mathrm{E}^{2/p}(\sigma_s^{(n)})^p &\leq \sum_{\ell=0}^{s+n} \gamma^\ell \mathrm{E}^{2/p}|Q(a+X_{s-\ell}^{(n)}|^p \\ &\leq \sum_{\ell=0}^{s+n} \gamma^\ell \mathrm{E}^{2/p}|c_1^2+c_2^2(a+X_{s-\ell}^{(n)})^2|^{p/2} \\ &\leq C + c_3^2 \sum_{\ell=0}^{n+s} \gamma^\ell \mathrm{E}^{2/p}(X_{s-\ell}^{(n)})^p, \end{aligned}$$

where $c_3 > c_2 > \mathrm{Lip}_Q$ and $c_3 - \mathrm{Lip}_Q > 0$ can be arbitrarily small. Particularly, for any fixed $T \in \mathbb{Z}$,

$$\sup_{-n\leq s<T} \mathrm{E}^{2/p}(\sigma_s^{(n)})^p \leq \frac{c_3^2}{(1-\gamma)} \sup_{-n\leq s<T} \mathrm{E}^{2/p}(X_s^{(n)})^p + C.$$

Substituting the last bound into inequality (5.28), we obtain

$$\sup_{-n\leq t<T} \mathrm{E}^{2/p}(X_t^{(n)})^p \leq \frac{\widetilde{\beta}^{2/p}}{(1-\gamma)} \sup_{-n\leq s<T} \mathrm{E}^{2/p}(X_s^{(n)})^p + C. \tag{5.30}$$

Relations (5.30) and (5.29) imply

$$\sup_{-n\leq t<T} \mathrm{E}^{2/p}(X_t^{(n)})^p \leq \frac{C}{1-\frac{\widetilde{\beta}^{2/p}}{(1-\gamma)}} < \infty$$

proving (5.25) as well as the theorem. □

Example 5.1. (Asymmetric GARCH(1,1)). The asymmetric GARCH(1,1) model of Engle [25] corresponds to

$$\sigma_t^2 = c^2 + (a+br_{t-1})^2 + \gamma\sigma_{t-1}^2, \tag{5.31}$$

or

$$\sigma_t^2 = \theta + \psi r_{t-1} + a_{11} r_{t-1}^2 + \delta \sigma_{t-1}^2, \tag{5.32}$$

in the parametrization of Sentana ([66], (5)), with parameters in equations (5.31) and (5.32) related by

$$\theta = c^2 + a^2, \quad \delta = \gamma, \quad \psi = 2ab, \quad a_{11} = b^2. \tag{5.33}$$

Under the condition that $\{\zeta_t = r_t/\sigma_t\}$ are standardized i.i.d., a stationary asymmetric GARCH(1,1) (or GQARCH(1,1) in the terminology of Sentana [66]) process $\{r_t, t \in \mathbb{Z}\}$ with finite variance and $a \neq 0$ exists if and only if $B_{2,\gamma} = b^2/(1-\gamma) < 1$, or

$$b^2 + \gamma < 1, \tag{5.34}$$

see Theorem 5.1 (ii). Condition (5.34) agrees with condition $a_{11} + \delta < 1$ for covariance stationarity in Sentana [66]. Under the assumptions that the distribution of ζ_t is symmetric and $\mu_4 = \mathrm{E}\zeta_t^4 < \infty$, Sentana [66] provides a sufficient condition for finiteness of $\mathrm{E}r_t^4$ together with explicit formula

$$\mathrm{E}r_t^4 \;=\; \frac{\mu_4 \theta[\theta(1 + a_{11} + \delta) + \psi^2]}{(1 - a_{11}^2 \mu_4 - 2a_{11}\delta - \delta^2)(1 - a_{11} - \delta)}. \tag{5.35}$$

The sufficient condition of Sentana [66] for $\mathrm{E}r_t^4 < \infty$ is $\mu_4 a_{11}^2 + 2a_{11}\delta + \delta^2 < 1$, which translates to

$$\mu_4 b^4 + 2b^2\gamma + \gamma^2 < 1, \tag{5.36}$$

in terms of the parameters of (5.31). Condition (5.36) seems weaker than the sufficient condition $\mu_4 b^4 + 6b^2 < (1-\gamma)^2$ of Theorem 5.2 for the existence of the L^4-solution of (5.31).

Following the approach of Doukhan *et al.* [22], below we find explicitly the covariance function $\rho(t) := \mathrm{cov}(r_0^2, r_t^2)$, including the expression in (5.35), for the stationary solution of the asymmetric GARCH(1,1) in (5.31). The approach of Doukhan *et al.* [22] is based on derivation and solution of linear equations for moment functions $m_2 := \mathrm{E}r_t^2$, $m_3(t) := \mathrm{E}r_t^2 r_0$ and $m_4(t) := \mathrm{E}r_t^2 r_0^2$. Assume that $\mu_3 = \mathrm{E}\zeta_0^3 = 0$, or $\mathrm{E}r_t^3 = 0$. We can write the following moment equations:

$$\begin{aligned} m_2 &= (c^2+a^2)/(1-b^2-\gamma), \quad m_3(0) = 0, \\ m_3(1) &= \sum_{\ell=0}^{\infty} \gamma^\ell \mathrm{E}(c^2+a^2+2abr_{-\ell}+b^2r_{-\ell}^2)r_0 = 2abm_2, \\ m_3(t) &= \sum_{\ell=0}^{\infty} \gamma^\ell \mathrm{E}(c^2+a^2+2abr_{t-\ell-1}+b^2r_{t-\ell-1}^2)r_0 \\ &= 2abm_2\gamma^{t-1} + b^2\sum_{\ell=0}^{t-2}\gamma^\ell m_3(t-\ell-1), \quad t \geq 2. \end{aligned} \tag{5.37}$$

From equations above, one can show by induction that

$$m_3(t) = 2abm_2(\gamma+b^2)^{t-1}, \quad t \geq 1.$$

Similarly,

$$\begin{aligned} m_4(0) &= \mu_4 \mathrm{E}((c^2+a^2)+2abr_0+b^2r_0^2+\gamma\sigma_0^2)^2 \\ &= \mu_4\left((c^2+a^2)^2+(2ab)^2m_2+b^4m_4(0)+\right. \\ &\quad \left.+2(c^2+a^2)(b^2+\gamma)m_2+(2b^2\gamma+\gamma^2)m_4(0)/\mu_4\right), \\ m_4(t) &= \sum_{\ell=0}^{\infty}\gamma^\ell \mathrm{E}(c^2+a^2+2abr_{t-\ell-1}+b^2r_{t-\ell-1}^2)r_0^2 \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\ell=0}^{\infty} \gamma^{\ell}(c^2+a^2)m_2 + b^2 \sum_{\ell=0}^{\infty} \gamma^{\ell} m_4(|t-\ell-1|) + \\
&\quad +2ab \sum_{\ell=t}^{\infty} \gamma^{\ell} m_3(\ell-t+1), \quad t \geq 1.
\end{aligned}
$$

Using

$$
\begin{aligned}
2ab \sum_{\ell=t}^{\infty} \gamma^{\ell} m_3(\ell-t+1) &= 4a^2b^2m_2 \sum_{\ell=t}^{\infty} \gamma^{\ell}(\gamma+b^2)^{\ell-t} \\
&= 4a^2b^2m_2\gamma^t/(1-\gamma(\gamma+b^2)),
\end{aligned}
$$

and $\rho(t) = m_4(t) - m_2^2$, we obtain the system of equations

$$
\begin{aligned}
\rho(0) &= m_4(0) - m_2^2, \\
\rho(t) &= b^2 \sum_{\ell=0}^{\infty} \gamma^{\ell} \rho(|t-\ell-1|) + 4a^2b^2m_2\gamma^t/(1-\gamma(\gamma+b^2)) \\
&= b^2 \sum_{\ell=0}^{t-2} \gamma^{\ell} \rho(t-\ell-1) + C\gamma^{t-1}, \quad t \geq 1, \qquad (5.38)
\end{aligned}
$$

where $C := b^2 \sum_{\ell=1}^{\infty} \gamma^{\ell} \rho(\ell) + (m_4(0) - m_2^2)b^2 + 4a^2b^2m_2\gamma/(1-\gamma(\gamma+b^2))$ is some constant independent of t, and

$$
m_4(0) = \frac{\mu_4 m_2}{1 - b^4\mu_4 - (2b^2\gamma + \gamma^2)} \Big((c^2+a^2)(1+b^2+\gamma) + (2ab)^2 \Big). \qquad (5.39)
$$

Note that the expression above coincides with (5.35) given that the relations in (5.33) hold.

Since the equation in (5.38) is analogous to (5.37), the solution to (5.38) is $\rho(t) = C(\gamma+b^2)^{t-1}, t \geq 1$. In order to find C, we combine $\rho(t) =$

$C(\gamma + b^2)^{t-1}$ and the expression for C to obtain the equation

$$C = Cb^2\gamma/(1 - \gamma(\gamma + b^2)) + (m_4(0) - m_2^2)b^2 + 4a^2b^2m_2\gamma/(1 - \gamma(\gamma + b^2)).$$

Now C can be expressed as

$$C = b^2 \frac{(m_4(0) - m_2^2)(1 - \gamma(\gamma + b^2)) + 4a^2m_2\gamma}{1 - \gamma(\gamma + 2b^2)},$$

together with equation (5.39) and $\rho(t) = C(\gamma + b^2)^{t-1}, t \geq 1,$ giving explicitly the covariances of process $\{r_t^2, t \in \mathbb{Z}\}$.

5.3 Long memory

The present section studies long memory properties of the generalized quadratic ARCH model in (5.4) corresponding to $Q(x) = \sqrt{c^2 + x^2}$ of equation (5.2), that is,

$$r_t \;=\; \zeta_t \sqrt{\sum_{\ell=0}^{\infty} \gamma^\ell \left(c^2 + \big(a + \sum_{s<t-\ell} b_{t-\ell-s} r_s\big)^2 \right)}, \qquad t \in \mathbb{Z}, \tag{5.40}$$

where $0 \leq \gamma < 1, a \neq 0, c$ are real parameters, $\{\zeta_t, t \in \mathbb{Z}\}$ are standardized i.i.d. random variables with zero mean and unit variance, and $b_j, j \geq 1$, are real numbers satisfying

$$b_j \;\sim\; \beta j^{d-1} \;\; (\exists\, 0 < d < 1/2,\; \beta > 0). \tag{5.41}$$

The main result of this section is Theorem 5.3, which shows that, under some additional conditions, the squared process $\{r_t^2, t \in \mathbb{Z}\}$ of equation (5.40) has similar long memory properties as in the case of the LARCH model (see Giraitis *et al.* [32], Theorem 2.2). Theorem 5.3 extends the result of Doukhan *et al.* ([22], Theorem 3) to the case $\gamma > 0$. In Theorem 5.3 and below, $0 \le \gamma < 1$, $B_2 = \sum_{j=1}^{\infty} b_j^2$ and $B(\cdot, \cdot)$ is a beta function.

Theorem 5.3. *Let $\{r_t, t \in \mathbb{Z}\}$ be a stationary L^2-solution of (5.40)–(5.41). Assume in addition that $\mu_4 = \mathrm{E}[\zeta_0^4] < \infty$, and $\mathrm{E}[r_t^4] < \infty$. Then*

$$\mathrm{cov}(r_0^2, r_t^2) \sim \kappa_1^2 t^{2d-1}, \qquad t \to \infty, \tag{5.42}$$

where $\kappa_1^2 := \big(\frac{2a\beta}{1-\gamma-B_2}\big)^2 B(d, 1-2d) \mathrm{E} r_0^2$. Moreover,

$$n^{-d-1/2} \sum_{t=1}^{[n\tau]} (r_t^2 - \mathrm{E} r_t^2) \to_{D[0,1]} \kappa_2 B_{d+(1/2)}(\tau), \qquad n \to \infty, \tag{5.43}$$

where $B_{d+(1/2)}$ is a fractional Brownian motion with the Hurst parameter $H = d + (1/2) \in (1/2, 1)$ and $\kappa_2^2 := \kappa_1^2/(d(1+2d))$.

To prove Theorem 5.3, we need the following two facts.

Lemma 5.1. (Doukhan *et al.* [22], Lemma 1) *For $\alpha_j \ge 0$, $j = 1, 2, \ldots$, denote*

$$A_k := \alpha_k + \sum_{0<p<k} \sum_{0<i_1<\cdots<i_p<k} \alpha_{i_1}\alpha_{i_2-i_1} \cdots \alpha_{i_p-i_{p-1}}\alpha_{k-i_p}, \quad k = 1, 2, \ldots .$$

Assume that $\sum_{j=1}^{\infty} \alpha_j < 1$ and

$$\alpha_j \le c\, j^{-\gamma}, \qquad (\exists\, c > 0,\ \gamma > 1).$$

Then there exists $C > 0$ such that for any $k \ge 1$

$$A_k \le C k^{-\gamma}.$$

Lemma 5.2. *Assume that* $0 \le \beta < 1$ *and* $\alpha_j \sim cj^{-\gamma}$ $(\exists\, \gamma > 0,\ c > 0)$. *Then*

$$\alpha_{t,\beta} := \sum_{j=0}^{t-1} \beta^j \alpha_{t-j} \ \sim\ \frac{c}{1-\beta}\, t^{-\gamma}, \qquad t \to \infty.$$

Proof. It suffices to show that the difference $D_t := \alpha_{t,\beta} - \alpha_t/(1-\beta)$ decays faster than α_t, in other words, that

$$D_t \ = \ \sum_{j=0}^{t-1} \beta^j (\alpha_t - \alpha_{t-j}) - \sum_{j=t}^{\infty} \beta^j \alpha_{t-j} = o(t^{-\gamma}).$$

Clearly, $\sum_{t/2<j<t} \beta^j(\alpha_t - \alpha_{t-j}) = O(\beta^{t/2}) = o(t^{-\gamma})$, $\sum_{j=t}^{\infty} \beta^j \alpha_{t-j} = O(\beta^t) = o(t^{-\gamma})$. Relation $\sum_{0\le j\le t/2} \beta^j(\alpha_t - \alpha_{t-j}) = o(t^{-\gamma})$ follows by the dominated convergence theorem since $\sup_{0\le j\le t/2} |\alpha_t - \alpha_{t-j}| t^{\gamma} \le C$ and $|\alpha_t - \alpha_{t-j}| t^{\gamma} \to 0$ for any fixed $j \ge 0$. $\square$

Proof of Theorem 5.3. We use the idea of the proof of Theorem 3 by Doukhan *et al.* [22]. Denote

$$\begin{aligned} b_{t,\gamma} \ &:= \ \sum_{j=0}^{t-1} \gamma^j b_{t-j}, \qquad & \tilde b^2_{t,\gamma} \ &:= \ \sum_{j=0}^{t-1} \gamma^j b^2_{t-j}, \qquad & t \ge 1, \\ X_t \ &:= \ \sum_{s<t} b_{t-s} r_s, \qquad & X_{t,\gamma} \ &:= \ \sum_{s<t} b_{t-s,\gamma} r_s, \qquad & t \in \mathbb{Z}. \end{aligned} \tag{5.44}$$

By the definition of r_t in (5.40), we have the following decomposition (c.f. Doukhan *et al.* [22], (6.11))

$$(r_t^2 - \mathrm{E} r_t^2) - \sum_{s<t} \tilde b^2_{t-s,\gamma} (r_s^2 - \mathrm{E} r_s^2) \ = \ 2a X_{t,\gamma} + U_t + V_{t,\gamma} \ =:\ \xi_t, \tag{5.45}$$

where $X_{t,\gamma}$ is the main term and the "remainder terms" U_t and $V_{t,\gamma}$ are given by

$$U_t := (\zeta_t^2 - \mathrm{E}\zeta_t^2)\sigma_t^2, \qquad V_{t,\gamma} := \sum_{\ell=0}^{\infty} \gamma^\ell V_{t-\ell}, \tag{5.46}$$

$$V_t := 2 \sum_{s_2<s_1<t} b_{t-s_1} b_{t-s_2} r_{s_1} r_{s_2}. \tag{5.47}$$

Using the identity $V_t = (X_t^2 - \mathrm{E}X_t^2) - \sum_{s<t} b_{t-2}^2 (r_t^2 - \mathrm{E}r_t^2)$, the convergence in L^2 of the series on the r.h.s. of equation (5.47) follows as in Doukhan *et al.* [22] (6.12). Hence, the series for $V_{t,\gamma}$ in (5.46) also converges in L^2.

Let us prove that

$$\mathrm{cov}(\xi_0, \xi_t) \sim 4a^2 \mathrm{cov}(X_{0,\gamma}, X_{t,\gamma}) \sim 4a^2 \lambda_1^2 t^{2d-1}, \ t \to \infty, \tag{5.48}$$

where $\lambda_1^2 = \beta^2/(1-\gamma)^2 B(d, 1-2d)$. The second relation in (5.48) follows from $b_{t,\gamma} \sim (\beta/(1-\gamma))t^{d-1}, t \to \infty$, see Lemma 5.1, and the fact that $X_{t,\gamma} = \sum_{s<t} b_{t-s,\gamma} r_s$ is a moving average in stationary uncorrelated innovations $\{r_s\}$. Since $\{U_t\}$ is also an uncorrelated sequence, $\mathrm{cov}(\xi_0, U_t) = 0 \, (t \geq 1)$, and the first relation in (5.48) is a consequence of

$$\mathrm{E}[U_0 X_{t,\gamma}] + \mathrm{E}[U_0 V_{t,\gamma}] = o(t^{2d-1}), \tag{5.49}$$

$$\mathrm{E}[X_{0,\gamma} V_{t,\gamma}] + E[V_{0,\gamma}(X_{t,\gamma} + V_{t,\gamma})] = o(t^{2d-1}). \tag{5.50}$$

We have

$$\mathrm{E}[U_0 X_{t,\gamma}] = b_{t,\gamma} \mathrm{E}[U_0 r_0] = O(t^{d-1}) = o(t^{2d-1})$$

and

$$\mathrm{E}[U_0 V_{t,\gamma}] = 2b_{t,\gamma} D_t = O(t^{d-1}) = o(t^{2d-1}),$$

where

$$|D_t| := |\mathrm{E}[U_0 r_0 \sum_{s<0} b_{t-s} r_s]| \le \mathrm{E}U_0^2 (\mathrm{E}r_0^4)^{1/2} (\mathrm{E}(\sum_{s<0} b_{t-s} r_s)^4)^{1/2} \le C$$

follows from Rosenthal's inequality in (5.11) since

$$\mathrm{E}\Big(\sum_{s<0} b_{t-s} r_s\Big)^4 \le K_4 \mathrm{E}r_0^4 \Big(\sum_{s<0} b_{t-s}^2\Big)^2 \le C.$$

This proves (5.49). The proof of (5.50) is analogous to that of Doukhan *et al.* [22] (6.13)–(6.14) and is omitted.

Next, let us prove (5.42). Recall the definition of $\tilde{b}_{j,\gamma}^2$ in (5.44). From the decomposition (5.45), we obtain

$$r_t^2 - \mathrm{E}r_t^2 \;=\; \sum_{i=0}^{\infty} \phi_{i,\gamma} \xi_{t-i}, \qquad t \in \mathbb{Z}, \tag{5.51}$$

where $\phi_{j,\gamma} \ge 0, j \ge 0$, are the coefficients of the power series

$$\Phi_\gamma(z) := \sum_{j=0}^{\infty} \phi_{j,\gamma} z^j = (1 - \sum_{j=1}^{\infty} \tilde{b}_{j,\gamma}^2 z^j)^{-1}, \; z \in \mathbb{C}, \; |z| < 1,$$

given by $\phi_{0,\gamma} := 1$,

$$\phi_{j,\gamma} \;:=\; \tilde{b}_{j,\gamma}^2 + \sum_{0<k<j} \sum_{0<s_1<\cdots<s_k<j} \tilde{b}_{s_1,\gamma}^2 \cdots \tilde{b}_{s_k - s_{k-1},\gamma}^2 \tilde{b}_{j-s_k,\gamma}^2, \quad j \ge 1.$$

From (5.41) and Lemmas 5.1 and 5.2, we infer that

$$\phi_{t,\gamma} = O(t^{2d-2}), \qquad t \to \infty, \tag{5.52}$$

in particular, $\Phi_\gamma(1) = \sum_{t=0}^{\infty} \phi_{t,\gamma} = (1-\gamma)/(1-\gamma-B_2) < \infty$ and the r.h.s.

of equation (5.51) is well defined. Relations (5.51) and (5.52) imply that

$$\operatorname{cov}(r_t^2, r_0^2) = \sum_{i,j=0}^{\infty} \phi_i \phi_j \operatorname{cov}(\xi_{t-i}, \xi_{-j}) \;\sim\; \Phi_\gamma^2(1) \operatorname{cov}(\xi_t, \xi_0), \quad t \to \infty, \quad (5.53)$$

see Doukhan *et al.* [22], (6.20). Now, (5.42) follows from relations (5.53) and (5.48). The invariance principle in (5.43) follows similarly as in the proof of Theorem 3 by Doukhan *et al.* [22], from (5.51), (5.48) and

$$n^{-d-1/2} \sum_{t=1}^{[n\tau]} X_{t,\gamma} \to_{D[0,1]} \lambda_2 W_{d+(1/2)}(\tau), \quad \lambda_2^2 = \lambda_1^2/d(1+2d),$$

the last fact being a consequence of a general result of Abadir *er al.* [1]. Theorem 5.3 is proved. □

5.4 Leverage

For the conditionally heteroscedastic model in (5.40) with $\mathrm{E}\zeta_t = \mathrm{E}\zeta_t^3 = 0, \mathrm{E}\zeta_t^2 = 1$, consider the leverage function $h_t = \operatorname{cov}(\sigma_t^2, r_0) = \mathrm{E}r_t^2 r_0,\ t \geq 1$. Following Giraitis *et al.* [33], and Doukhan *et al.* [22], we say that $\{r_t, t \in \mathbb{Z}\}$, in equation (5.40) *has leverage of order* $k \geq 1$ (denoted by $\{r_t\} \in \ell(k)$) if

$$h_j < 0, \qquad 1 \leq j \leq k.$$

The study by Doukhan *et al.* [22] of leverage for model (5.40) with $\gamma = 0$, that is,

$$r_t \;=\; \zeta_t \sqrt{c^2 + \left(a + \sum_{s<t} b_{t-s} r_s\right)^2}, \qquad t \in \mathbb{Z},$$

was based on a linear equation for the leverage function:

$$h_t = 2am_2b_t + \sum_{0<i<t} b_i^2 h_{t-i} + 2b_t \sum_{i>0} b_{i+t}h_i, \qquad t \geq 1,$$

where $m_2 = \mathrm{E}r_0^2$. A similar equation (5.55) for the leverage function can be derived for model (5.40) in the general case $0 \leq \gamma < 1$. Namely, using $\mathrm{E}r_s = 0$, $\mathrm{E}r_s r_0 = m_2 \mathbf{1}(s = 0)$, $\mathrm{E}r_s^2 r_0 = 0\,(s \leq 0)$, $\mathrm{E}r_0 r_{s_1} r_{s_2} = \mathbf{1}(s_1 = 0)h_{-s_2}\,(s_2 < s_1)$ as in Doukhan *et al.* [22] we have that

$$\begin{aligned} h_t &= \mathrm{E}r_t^2 r_0 = \sum_{\ell=0}^{t-1} \gamma^\ell \mathrm{E}\big[(c^2 + (a + \sum_{s<t-\ell} b_{t-\ell-s}r_s)^2)r_0\big] \\ &= \sum_{\ell=0}^{t-1} \gamma^\ell \big(2am_2b_{t-\ell} + \sum_{s<t-\ell} b_{t-\ell-s}^2 \mathrm{E}[r_s^2 r_0]\big) + \qquad (5.54) \\ &\quad +2\sum_{\ell=0}^{t-1} \gamma^\ell \sum_{s_2<s_1<t-\ell} b_{t-\ell-s_1}b_{t-\ell-s_2}\mathrm{E}[r_{s_1}r_{s_2}r_0] \\ &= 2am_2b_{t,\gamma} + \sum_{0<i<t} h_i \tilde{b}_{t-i,\gamma}^2 + 2\sum_{i>0} h_i w_{i,t,\gamma}, \qquad (5.55) \end{aligned}$$

where $b_{t,\gamma}, \tilde{b}_{t,\gamma}^2$ are defined in (5.44) and $w_{i,t,\gamma} := \sum_{\ell=0}^{t-1} \gamma^\ell b_{t-\ell} b_{i+t-\ell}$.

Proposition 5.3. *Let $\{r_t, t \in \mathbb{Z}\}$ be a stationary L^2-solution of equation (5.40) with $\mathrm{E}|r_0|^3 < \infty$, $|\mu|_3 < \infty$. Assume, in addition, that $B_{2,\gamma} < 1/5$, $\mu_3 = \mathrm{E}\zeta_0^3 = 0$. Then for any fixed k such that $1 \leq k \leq \infty$:*

(i) *if $ab_1 < 0$, $ab_j \leq 0, j = 2, \dots, k$, then $\{r_t\} \in \ell(k)$;*

(ii) *if $ab_1 > 0$, $ab_j \geq 0, j = 2, \dots, k$, then $h_j > 0$, for $j = 1, \dots, k$.*

Proof. Let us prove that

$$\|h\| := \left(\sum_{t=1}^{\infty} h_t^2\right)^{1/2} \leq \frac{2|a|m_2 B_2^{1/2}}{(1-\gamma)(1-3B_{2,\gamma})}. \qquad (5.56)$$

Let $|b|_{t,\gamma} := \sum_{\ell=0}^{t-1} \gamma^\ell |b_{t-\ell}|$. By Minkowski's inequality,

$$\Big(\sum_{i=1}^\infty w_{i,t,\gamma}^2\Big)^{1/2} \ \le\ \sum_{\ell=0}^{t-1} \gamma^\ell |b_{t-\ell}| \Big(\sum_{i=1}^\infty b_{i+t-\ell}^2\Big)^{1/2} \ \le\ |b|_{t,\gamma} B_2^{1/2}, \quad (5.57)$$

and, therefore, $|\sum_{i=1}^\infty h_i w_{i,t,\gamma}| \le \|h\| B_2^{1/2} |b|_{t,\gamma}$. Moreover,

$$\Big(\sum_{t=1}^\infty b_{t,\gamma}^2\Big)^{1/2} \le B_2^{1/2}/(1-\gamma), \quad \Big(\sum_{t=1}^\infty |b|_{t,\gamma}^2\Big)^{1/2} \le B_2^{1/2}/(1-\gamma)$$

and $\big(\sum_{t=1}^\infty (\sum_{0<i<t} h_i \tilde b_{t-i,\gamma}^2)^2\big)^{1/2} \le \|h\| B_{2,\gamma} = \|h\| B_2/(1-\gamma)$. The above inequalities together with (5.55) imply

$$\|h\| \ \le\ 2|a| m_2 B_2^{1/2}/(1-\gamma) + \|h\| B_2/(1-\gamma) + 2\|h\| B_2^{1/2} B_2^{1/2}/(1-\gamma),$$

proving (5.56).

Using (5.55) and (5.56), statements (i) and (ii) can be proved by induction on $k \ge 1$ similar to Doukhan *et al.* [22]. Since $w_{i,1,\gamma} = b_1 b_{i+1}$ and $b_{1,\gamma} = b_1$, equation (5.55) yields

$$h_1 \ =\ 2am_2 b_{1,\gamma} + 2\sum_{i>0} w_{i,1,\gamma} h_i = 2b_1\big(am_2 + \sum_{i>0} h_i b_{i+1}\big). \quad (5.58)$$

According to (5.56), the last sum in (5.58) does not exceed

$$\Big|\sum_{i>0} h_i b_{i+1}\Big| \le \|h\| B_2^{1/2} \le 2|a| m_2 B_{2,\gamma}/(1-3B_{2,\gamma}) < |a| m_2$$

provided $B_{2,\gamma} < 1/5$. Hence, (5.58) implies sgn$(h_1) =$ sgn(ab_1), or statements (i) and (ii) for $k = 1$.

Let us prove the induction step $k-1 \to k$ in (i). Assume first that

$a > 0, b_1 < 0, b_2 \leq 0, \cdots, b_{k-1} \leq 0$. Then $h_1 < 0, h_2 < 0, \cdots, h_{k-1} < 0$ by the inductive assumption. By (5.55),

$$h_k = 2\Big(am_2 b_{k,\gamma} + \sum_{i>0} h_i w_{i,k,\gamma}\Big) + \sum_{0<i<k} \tilde{b}_{i,\gamma}^2 h_{k-i},$$

where $\sum_{0<i<k} \tilde{b}_{i,\gamma}^2 h_{k-i} < 0$ and $|\sum_{i>0} h_i w_{i,k,\gamma}| \leq \|h\| B_2^{1/2} |b|_{k,\gamma} < am_2 |b|_{k,\gamma}$ according to (5.56), (5.57). Since $b_{k,\gamma} < 0$ and $|b|_{k,\gamma} = |b_{k,\gamma}|$, this implies $am_2 b_{k,\gamma} + \sum_{i>0} h_i w_{i,k,\gamma} \leq 0$, or $h_k < 0$. The remaining cases in (i)–(ii) follow analogously. □

Chapter 6

Conclusions

In this last chapter, we summarize the main results of the dissertation:

- in this dissertation we showed that FIGARCH and IARCH(∞) equations with zero intercept may have a nontrivial covariance stationary solution with long memory;
- we provided a complete answer to the long standing conjecture of Ding and Granger ([20], 1996) about the existence of the Long Memory ARCH model;
- we introduced and investigated a new class of long memory integrated AR(p, d, q) processes and showed that their autocovariance can be modeled easily at low lags without a significant effect on the long memory behavior, this being a major advantage over classical ARFIMA models;
- we also obtained necessary and sufficient conditions for the existence of stationary integrated AR(∞) processes with finite variance and proved that such processes always have long memory;

- we studied the five-parametric QML estimation for a quadratic ARCH process with long memory and strictly positive conditional variance. Several QML estimators of unknown parameter $\theta_0 \in \mathbb{R}^5$ of our model were discussed, in particular, an estimator depending on observations from the infinite past, and a class of estimators depending only on observations from the finite past. We proved consistency and asymptotic normality of these estimators;

- a simulation study of the empirical MSE of QML estimation was included. In the simulation experiment, we studied the empirical performance of a more realistic version of the estimator and showed that the empirical RMSEs of this estimator show a good agreement with theoretical standard deviations;

- we studied the existence and properties of a stationary solution of the ARCH-type equation $r_t = \zeta_t \sigma_t$, where the conditional variance satisfies $\sigma_t^2 = Q^2\big(a + \sum_{j=1}^{\infty} b_j r_{t-j}\big) + \gamma \sigma_{t-1}^2$ with a Lipschitz function $Q(x)$ and real parameters a, γ, b_j. We obtained conditions for the existence of a stationary solution, and, in particular, when Q is the square root of a quadratic polynomial, we proved that r_t can exhibit a leverage effect and long memory.

Bibliography

[1] K. M. Abadir, W. Distaso, L. Giraitis, and H. L. Koul. Asymptotic normality for weighted sums of linear processes. *Econometric Theory*, 30:252–284, 2014.

[2] B. D. O. Andersen and J. B. Moore. *Optimal Filtering*. Prentice Hall, Inc., Englewood Cliffs, N.J., 1979.

[3] R. T. Baillie, T. Bollerslev, and H. O. Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 74:3–30, 1996.

[4] J. Beran. *Statistics for long-memory processes*. Chapman and Hall, New York, 1994.

[5] J. Beran and M. Schützner. On approximate pseudo-maximum likelihood estimation for LARCH-processes. *Bernoulli*, 15:1057–1081, 2009.

[6] I. Berkes and L. Horváth. The rate of consistency of the quasi-maximum likelihood estimator. *Statistics and Probability Letters*, 61: 133–143, 2003.

[7] I. Berkes and L. Horváth. The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics*, 32:633–655, 2004.

[8] I. Berkes, L. Horváth, and P. S. Kokozska. GARCH processes: struc-

ture and estimation. *Bernoulli*, 9:201–227, 2003.

[9] P. Billingsley. *Convergence of Probability Measures*. New York: Wiley, 1968.

[10] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[11] A. Bose and K. Mukherjee. Estimating the ARCH parameters by solving linear equations. *Journal of Time Series Analysis*, 24:127–136, 2003.

[12] P. Bougerol and N. Picard. Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 52:115–127, 1992.

[13] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods (2nd edition)*. Springer Series in Statistics. Springer-Verlag, New York., 1991.

[14] DL. Burkholder. Distribution functions inequalities for martingales. *Annals of Probability*, 1:19–42, 1973.

[15] V. Kazakevičius and R. Leipus. On stationarity in the ARCH(∞) model. *Econometric Theory*, 18:1–16, 2002.

[16] V. Kazakevičius and R. Leipus. A new theorem on existence of invariant distributions with applications to ARCH processes. *Journal of Applied Probability*, 40:147–162, 2003.

[17] D. R. Cox. Long-range dependence: a review. In *Statistics: An Appraisal. Proc. 50th Anniversary Conference*, pages 55–74. Iowa State University Press, 2009.

[18] J. Davidson. Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economic Statistics*, 22:16–29, 2004.

[19] R. B. Davies and D. S. Harte. Tests for Hurst effect. *Biometrica*, 74: 95–102, 1987.

[20] Z. Ding and C. W. J. Granger. Modelling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 73: 185–215, 1996.

[21] R. Douc, F. Roueff, and P. Soulier. On the existence of some ARCH(∞) processes. *Stochastic Processes and their Applications*, 118:755–761, 2008.

[22] P. Doukhan, I. Grublytė, and D. Surgailis. A nonlinear model for long memory conditional heteroscedasticity. *Lithuanian Mathematical Journal*, 56:164–188, 2016.

[23] A. Osękowski. A note on Burkholder-Rosenthal inequality. *Bulletin of the Polish Academy Sciencies. Mathematics*, 60:177–185, 2012.

[24] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50: 987–1008, 1982.

[25] R. F. Engle. Stock volatility and the crash of '87. *Discussion. The Review of Financial Studies*, 3:103–106, 1990.

[26] C. Francq and JM. Zakoian. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10:605–637, 2004.

[27] C. Francq and JM. Zakoian. A tour in the asymptotic theory of GARCH estimation. In *Handbook of financial time series*, pages 85–109. Springer-Verlag Berlin Heidelberg, 2009.

[28] C. Francq and JM. Zakoian. *GARCH models: Structure, Statistical Inference and Financial Applications*. New York: Wiley, 2010.

[29] C. Francq and JM. Zakoian. Inconsistency of the MLE and inference based on weighted LS for LARCH models. *Journal of Econometrics*, 159:151–165, 2010.

[30] L. Giraitis and D. Surgailis. ARCH-type bilinear models with double long memory. *Stochastic Processes and their Applications*, 100:275–300, 2002.

[31] L. Giraitis, P. Kokoszka, and R. Leipus. Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory*, 16:3–22, 2000.

[32] L. Giraitis, P. M. Robinson, and D. Surgailis. A model for long memory conditional heteroscedasticity. *The Annals of Applied Probability*, 10:1002–1024, 2000.

[33] L. Giraitis, R. Leipus, P. M. Robinson, and D. Surgailis. LARCH, leverage and long memory. *Journal of Financial Econometrics*, 2:177–210, 2004.

[34] L. Giraitis, R. Leipus, and D. Surgailis. Recent advances in ARCH modelling. In *Long Memory in Economics*, pages 3–38. Springer-Verlag, 2007.

[35] L. Giraitis, R. Leipus, and D. Surgailis. ARCH(∞) models and long memory properties. In *Handbook of financial time series*, pages 71–84. Springer-Verlag Berlin Heidelberg, 2009.

[36] L. Giraitis, H. L. Koul, and D. Surgailis. *Large sample inference for long memory processes*. Imperial College press, 2012. ISBN 978-1-84816-278-5.

[37] L. Giraitis, D. Surgailis, and A. Škarnulis. Stationary integrated ARCH(∞) and AR(∞) processes with finite variance. *Under revision in Econometric Theory*, 2017.

[38] C. W. J. Granger and A. P. Andersen. *An Introduction to Bilinear Time Series Models*. Vandenhoek and Ruprecht, Göttingen, 1978.

[39] C. W. J. Granger and Z. Ding. Some properties of absolute return: an

alternative measure of risk. *Annales d'Economie et de Statistique*, 40: 67–91, 1995.

[40] I. Grublytė and A. Škarnulis. A nonlinear model for long memory conditional heteroscedasticity. *Statistics*, 51:123–140, 2017.

[41] I. Grublytė, D. Surgailis, and A. Škarnulis. QMLE for quadratic ARCH model with long memory. *Journal of Time Series Analysis*, 2016. doi: 10.1111/jtsa.12227.

[42] P. Hall and Q. Yao. Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, 71:285–317, 2003.

[43] J. D. Hamilton. *Time series analysis*. Princeton University Press, Princeton, 1994.

[44] J. R. M. Hosking. Fractional differencing. *Biometrika*, 68:165–176, 1981.

[45] J. K. Hunter and B. Nachtergaele. *Applied Analysis*. World Scientific Publishing Co.Pte. Ltd., 2001.

[46] H. Hurst. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–808, 1951.

[47] H. Hurst. Methods of using long-term storage in reservoirs. *Proceedings of the Institution of Civil Engineers, Part I*, pages 519–577, 1955.

[48] I. A. Ibragimov and Yu. V. Linnik. *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen, 1971.

[49] P. Kokoszka and R. Leipus. Change-point estimation in ARCH models. *Bernoulli*, 6:513–539, 2000.

[50] D. Koulikov. Long memory ARCH(∞) models: Specification and quasi-maximum likelihood estimation. *Working Paper 163, Centre for Analytical Finance, Univ. Aarhus. www.cls.dk/caf/wp/wp-163.pdf*, 2003.

[51] R. Leipus and V. Kazakevičius. On stationarity in the ARCH(∞)

model. *Econometric Theory*, 18:1–16, 2002.

[52] M. Levine, S. Torres, and F. Viens. Estimation for the long-memory parameter in LARCH models, and fractional Brownian motion. *Statistical Inference for Stochastic Processes*, 12:221–250, 2009.

[53] O. Lieberman and P. C. B. Phillips. Refined inference on long memory in realized volatility. *Econometric Reviews*, 27:254–267, 2008.

[54] A.M. Lindner. Statioanrity, mixing, distributional properties and moments of GARCH(p,q)-processes. In *Handbook of financial time series*, pages 43–69. Springer-Verlag Berlin Heidelberg, 2009.

[55] B. Mandelbrot. Une classe de processus stochastiques homothetiques a soi; application a loi climatologique de h. e. hurst. *Comptes Rendus Academic Sciences Paris*, 240:3274–3277, 1965.

[56] B. Mandelbrot and J. Wallis. Noah, Joseph and operational hydrology. *Water Resources Research*, 4:909–918, 1968.

[57] T. Mikosch and C. Stărică. Is it really long memory we see in financial returns? In *Extremes and Integrated Risk Management, London*, pages 149–168. Risk Books, 2000.

[58] T. Mikosch and C. Stărică. Long-range dependence effects and ARCH modeling. In *Theory and Applications of Long-Range Dependence*, pages 439–459. Birkhäuser, 2003.

[59] P. Moran. On the range of cumulative sums. *Annals of the Institute of Statistical Mathematics*, 16:109–112, 1964.

[60] D. B. Nelson. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6:318–334, 1990.

[61] T. Subba Rao. On the theory of bilinear time series models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43:244–255, 1981.

[62] P. M. Robinson. Testing for strong serial correlation and dynamic

conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47:67–84, 1991.

[63] P. M. Robinson and P. Zaffaroni. Pseudo-maximum likelihood estimation of ARCH(∞) models. *The Annals of Statistics*, 34:1049–1074, 2006.

[64] HP. Rosenthal. On the subspaces of $l^p (p > 2)$ spanned by the sequences of independent random variables. *Israel Journal of Mathematics*, 8:273–303, 1970.

[65] W. Rudin. *Real and Complex Analysis*. McGraw-Hill Book Company, New York etc., 1987.

[66] E. Sentana. Quadratic ARCH models. *The Review of Economic Studies*, 3:77–102, 1995.

[67] W. Stout. *Almost sure convergence*. New York: Academic Press, 1974.

[68] L. Truquet. On a family of contrasts for parametric inference in degenerate ARCH models. *Econometric Theory*, 30:1165–1206, 2014.

[69] B. von Bahr and C-G. Esséen. Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *The Annals of Mathematical Statistics*, 36:299–303, 1965.

[70] P. Whittle. Estimation and information in stationary time series. *Arkiv för Matematik*, 2:423–443, 1953.