VILNIUS UNIVERSITY

IEVA GRUBLYTĖ

MODELLING OF NONLINEAR LONG MEMORY

Doctoral Dissertation
Physical Sciences, Mathematics (01P)

Vilnius, 2017

The scientific work was carried out in 2013-2017 at Vilnius University and University of Cergy-Pontoise (France).

Scientific Supervisor - Prof. Dr. Habil. Donatas Surgailis (Vilnius University, Physical sciences, Mathematics - 01P).
Scientific Advisor - Prof. Dr. Paul Doukhan (University of Cergy-Pontoise, Physical sciences, Mathematics - 01P).

VILNIAUS UNIVERSITETAS

IEVA GRUBLYTĖ

NETIESINIAI ILGOS ATMINTIES MODELIAI

Daktaro disertacija
Fiziniai mokslai, matematika (01P)

Vilnius, 2017

Disertacija rengta 2013-2017 metais Vilniaus Universitete ir Cergy-Pontoise Universitete (Prancūzija).

Mokslinis vadovas - prof. habil. dr. Donatas Surgailis (Vilniaus universitetas, fiziniai mokslai, matematika - 01P).
Mokslinis konsultantas - prof. dr. Paul Doukhan (Cergy-Pontoise universitetas, fiziniai mokslai, matematika - 01P).

Contents

Notations and Abbreviations

$\mathbb{N}$ - the set of natural numbers, $\mathbb{N} = \{1, 2, \dots\}$

$\mathbb{Z}$ - the set of natural numbers, $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$

$\mathbb{R}$ - the set of real numbers

$C, C(\dots)$ denote generic constants, possibly dependent on the variables into brackets, which may be different at different locations

$\mathrm{E}X$ denotes the mean of random variable X

$\mathrm{Var}(X)$ denotes the variance of random variable X

$\mathrm{Cov}(X, Y)$ denotes the covariance of random variables X, Y

$sign(\cdot)$ is a sign function

$\mathbf{1}_{(\cdot)}$, $\mathbf{1}(.)$ denote the indicator function

$x \wedge y$ denotes $min(x, y)$ for real numbers x, y

$x \vee y$ denotes $max(x, y)$ for real numbers x, y

L is a lag operator, i.e. $LX_t = X_{t-1}$

$B(\cdot, \cdot)$ is a beta function

$\Gamma(\cdot)$ is a gamma function

$B_H(t)$ denotes fractional Brownian motion where H is the Hurst parameter

$\to_{D[0,1]}$ denotes the weak convergence of random processes in the Skorohod space $D[0, 1]$

$\| \cdot \|_p := \mathrm{E}^{1/p} | \cdot |^p, p \geq 1$ denotes L_p norm

$\ell^\infty(\mathbb{R})$ denotes the space of all bounded functions on $\mathbb{R}$

i.i.d independent identically distributed

r.v random variable

a.s. almost surely

r.h.s right hand side

l.h.s left hand side

w.r.t with respect to

Chapter 1

Introduction

Long memory as an object of research

A discrete-time second-order stationary process $\{X_t, t \in \mathbb{Z}\}$ is called *long memory* if its covariance $\gamma(k) = \mathrm{Cov}(X_0, X_k)$ decays slowly with the lag in such a way that its absolute series diverges:

$$\sum_{k=1}^{\infty} |\gamma(k)| = \infty. \tag{1.1}$$

In the converse case when

$$\sum_{k=1}^{\infty} |\gamma(k)| < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \gamma(k) \neq 0 \tag{1.2}$$

the process $\{X_t\}$ is said to have *short memory*. *Negative memory* is defined as

$$\sum_{k=1}^{\infty} |\gamma(k)| < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \gamma(k) = 0. \tag{1.3}$$

Long memory processes have different properties from short memory (in particular, i.i.d.) processes. Long memory processes have been found to arise in a variety of physical and social sciences. See, e.g., the monographs Beran (1997), Doukhan et al. (2003), Giraitis et al. (2012), Beran et al. (2013) and the references therein.

Conditions (1.1)-(1.3) defining long, short and negative memory are very general. A useful asymptotic theory and statistical inference is possible if one specifies the rate of decay of $\gamma(k)$ at infinity. Particularly, (1.1) is often specified as

$$\gamma(k) = |k|^{-1+2d} L_\gamma(|k|), \quad |k| \geq 1, \qquad 0 < d < 1/2 \tag{1.4}$$

or

$$\gamma(k) \;\sim\; c_\gamma |k|^{-1+2d}, \quad k \to \infty, \quad 0 < d < 1/2, \quad c_\gamma > 0, \tag{1.5}$$

where $L_\gamma : [1, \infty) \to \mathbb{R}$ is a slowly varying function at infinity. Parameter $d \in (0, 1/2)$ in (1.4) and (1.5) is called the *long memory parameter* of $\{X_t\}$. It characterizes the intensity of long memory of the process $\{X_t\}$: when $d > 0$ is small the covariance function decays relatively fast and the intensity of long memory is small, while in the case when d is close to $1/2$ the covariance function decays very slowly since the exponent $-1+2d$ is almost zero, and the corresponding process $\{X_t\}$ has very strong memory.

Probably, the most important model of long memory processes is the *linear*, or moving average process

$$X_t \;=\; \sum_{s \le t} b_{t-s}\zeta_s, \qquad t \in \mathbb{Z}, \tag{1.6}$$

where $\{\zeta_s, s \in \mathbb{Z}\}$ is a standardized i.i.d. sequence, and the moving average coefficients b_j decay slowly so that $\sum_{j=0}^\infty |b_j| = \infty, \sum_{j=0}^\infty b_j^2 < \infty$. The last condition guarantees that the series in (1.6) converges in mean square and satisfies $\mathrm{E}X_t = 0, \mathrm{E}X_t^2 = \sum_{j=0}^\infty b_j^2 < \infty$. In the literature it is often assumed that the coefficients regularly decay as

$$b_j \sim \kappa j^{d-1}, \qquad j \to \infty \qquad (\exists\, \kappa > 0,\ 0 < d < 1/2). \tag{1.7}$$

Condition (1.7) guarantees (1.5), i.e. that

$$\gamma(k) \;=\; \sum_{j=0}^\infty b_j b_{k+j} \;\sim\; \kappa^2 B(d, 1-2d) k^{-1+2d}, \qquad k \to \infty \tag{1.8}$$

and hence $\sum_{k=1}^\infty |\gamma(k)| = \infty$. Thus, the parameter d in (1.7) is the long memory parameter of $\{X_t\}$ as defined in (1.5).

An important property of the linear process in (1.6)-(1.7) is the fact that its (normalized) partial sums process $S_n(\tau) := \sum_{j=1}^{[nt]} X_j$, $\tau \ge 0$ tends to a fractional Brownian motion (Davydov (1970)), viz.,

$$n^{-d-1/2} S_n(\tau) \;\to_{D[0,1]}\; \sigma(d) B_H(t), \tag{1.9}$$

where $H = d + \frac{1}{2}$ is the Hurst parameter, $\sigma(d)^2 := \kappa^2 B(d, 1-2d)/d(1+2d) > 0$ and $\to_{D[0,1]}$ denotes the weak convergence of random processes in the Skorohod space $D[0,1]$. By definition, fractional Brownian motion is a Gaussian process with zero mean and covariance function

$$\mathrm{E}B_H(s)B_H(t) = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}), \qquad t, s \geq 0. \tag{1.10}$$

Note that the normalization in (1.9) grows faster than the classical normalization $n^{1/2}$ in Donsker's invariance principle for weakly dependent summands, and the limit process B_H has *dependent increments* in contrast to the usual Brownian motion with independent increments. Fractional Brownian motion is H-self-similar and plays a very important role in many applications of stochastic processes. The above mentioned properties of the partial sums limit are characteristic to long memory although in general the covariance decay as in (1.5) does not imply a fractional Brownian motion limit of the partial sums process.

Motivation and aims of the thesis

It is well-known that the linear model (1.6) has its drawbacks and sometimes is not capable of incorporating empirical features ("stylized facts") of some observed time series. The "stylized facts" may include typical asymmetries, clusterings, and other nonlinearities which are often observed in financial data, together with long memory. A very important stylized fact of asset returns is *conditional heteroscedasticity*, or the property of the conditional variance

$$\mathrm{Var}[X_{t+1}|\mathcal{F}_t] = \mathrm{E}[(X_{t+1} - \mathrm{E}[X_{t+1}|\mathcal{F}_t])^2|\mathcal{F}_t]$$

being a random process and not a constant like in linear models. Here, $\mathcal{F}_t$ is the "historic" σ-field containing "all available information" and $\mathrm{E}[X_{t+1}|\mathcal{F}_t]$ is the best forecast of X_{t+1} given the "information" $\mathcal{F}_t$. For the linear process in (1.6) and $\mathcal{F}_t = \sigma\{\zeta_s, s \leq t\}$, the best forecast is $\mathrm{E}[X_{t+1}|\mathcal{F}_t] = b_1\zeta_t + b_2\zeta_{t-2} + \ldots$ and $X_{t+1} - \mathrm{E}[X_{t+1}|\mathcal{F}_t] = b_0\zeta_{t+1}$ is independent of $\mathcal{F}_t$ so that the conditional variance is constant: $\mathrm{Var}[X_{t+1}|\mathcal{F}_t] = b_0^2\mathrm{E}\zeta_t^2$, meaning that this model is conditionally homoscedastic. Therefore developing nonlinear models with long memory presents considerable interest.

Problems and main results

The principal goal of this thesis is to introduce *new nonlinear models with long memory* that could be used for modelling of financial returns and statistical inference.

1. *Projective stochastic equations (Chapter 3).*

The main goal is to introduce a new class of nonlinear processes which generalize the linear model in (1.6)-(1.7) and enjoy similar long memory properties to (1.8) and (1.9). For this, we define projective moving averages $\{X_t, t \in \mathbb{Z}\}$, where X_t is a Bernoulli shift written as a backward martingale transform of the innovation sequence. We introduce a new class of nonlinear stochastic equations for projective moving averages, termed *projective equations*, involving a (nonlinear) kernel Q and a linear combination of projections of X_t on "intermediate" lagged innovation subspaces with given coefficients $\alpha_i, \beta_{i,j}$. We obtain conditions for solvability of these equations. We also show that under certain conditions on kernel and coefficients, the solution exhibits covariance and distributional long memory, with fractional Brownian motion as the limit of the corresponding partial sums process. Results are presented in Chapter 3 and Grublytė and Surgailis (2014).

2. A nonlinear model for long memory conditional heteroscedasticity (Chapter 4).

The main goal is to introduce a new class of conditionally heteroscedastic processes that generalize some of the already known models and are able to model long memory and other stylized facts in certain cases. For this, we discuss a class of conditionally heteroscedastic time series models satisfying the ARCH-type equation $r_t = \zeta_t \sigma_t$, where ζ_t is a noise sequence and the conditional standard deviation σ_t is a nonlinear function Q depending on a linear combination of past values $r_s, s < t$ with coefficients b_j. We obtain the conditions for the existence of stationary solution r_t with finite p-th moment, $0 < p < \infty$. Weak dependence properties of r_t are studied, including the invariance principle for partial sums of Lipschitz functions of r_t. The case when Q is the square root of a quadratic polynomial corresponds to a quadratic ARCH (QARCH) model and is of special interest. We prove that in this case r_t can exhibit a leverage effect and long memory, in the sense that the squared process r_t^2 has long memory autocorrelation and its normalized partial sums process converges to a fractional Brownian motion. Analogous results are obtained for the generalized version of the model described above where the conditional variance satisfies an AR(1) equation, i.e. the volatily form includes the lagged volatilities from the past. We also obtain a new condition for the existence of higher moments of r_t which does not include the Rosenthal constant. A short simulation study showing the behavior of processes defined by this model is included. Results are presented in Chapter 4 and Doukhan et al. (2016), Grublytė and Škarnulis (2017).

3. Quasi-MLE for quadratic ARCH model with long memory (Chapter 5).

The goal is to provide the asymptotic results for quasi-maximum likelihood estimators in parametric version of long memory QARCH model introduced in Chapter 4 (also Doukhan et al. (2016), Grublytė and Škarnulis (2017)). Similarly as in Beran and Schützner (2009) we discuss several QML estimators: the estimator involving exact conditional variance depending on infinite past and its more realistic version where the volatilities depend only on finite number of returns from past. Under certain moment conditions we prove consistency and asymptotic normality of the corresponding QML estimators, including the estimator of long memory parameter $0 < d < 1/2$. Results are presented in Chapter 5 and Grublytė et al. (2017).

The novelty

New nonlinear models with long memory for modelling of financial returns are developed in this thesis. These processes are defined as stationary solutions of certain nonlinear stochastic difference equations involving a given i.i.d. "noise". Solvability of these equations is studied and covariance and distributional long memory is proved. Finally, for a particularly tractable nonlinear parametric model with long memory (GQARCH) consistency and asymptotic normality of quasi-ML estimators are proved.

The processes studied in the thesis are new and have not been investigated in a scientific literature before.

Methods

Many proofs in the thesis use the idea of projections (discussed in more detail in Chapter 3, Section 3.2). Besides that, other standard tools from probability theory, functional analysis, mathematical statistics and time series analysis were used.

Dissemination

The results were presented in the following conferences and seminars:

- 54th conference of Lithuanian Mathematical Society, Vilnius (Lithuania), June 19-20, 2013.
- 55th conference of Lithuanian Mathematical Society, Vilnius (Lithuania), June 26-27, 2014.
- 11th International Vilnius Conference on Probability Theory and Mathematical Statistics, Vilnius (Lithuania), June 30 - July 4, 2014.
- Séminaire SAMM: Statistique, Analyse et Modélisation Multidisciplinaire, Université Paris 1 Panthéon-Sorbonne, Paris (France), November 28, 2015.

- 8th annual SoFiE (The Society for Financial Econmetrics) conference, Aarhus (Denmark), June 23 - 26, 2015.
- Conference “Stochastic Processes”, Luminy (France), February 15-19, 2016.
- Seminar at “Workshop on Dependence”, Institut Henri Poincaré, Paris (France), September 27, 2016.

Publications

- I. Grublytė, D. Surgailis (2014). Projective stochastic equations and nonlinear long memory, Adv. in Appl. Probab., 46(4):1-22.
- P. Doukhan, I. Grublytė, D. Surgailis (2016). A nonlinear model for long memory conditional heteroscedasticity, Lith. Math. J. 56(2):164–188.
- I. Grublytė, A. Škarnulis (2017). A generalized nonlinear model for long memory conditional heteroscedasticity, Statistics 51(1):123-140.
- I. Grublytė, D. Surgailis, A. Škarnulis (2017). QMLE for quadratic ARCH model with long memory. J. Time Ser. Anal. 38(4):535–551.

Structure of thesis

The thesis consists of Introduction, State of the Art, three Chapters, Conclusions, two Appendixes and Bibliography. The review of aims and problems is given in Introduction. State of the Art presents an overview of the scientific work in this field. Chapter 3 introduces nonlinear processes defined through projective stochastic equations. Chapter 4 presents a very general class of nonlinear conditionally heteroscedastic models. A separate case of (G)QARCH model ((Generalized) Quadratic ARCH) is studied in more detail. Chapter 5 considers the estimation of parameters in generalized QARCH models using QML method. The results of the thesis are summarized in Conclusions.

Acknowledgements

I’m very grateful to my advisors prof. Donatas Surgailis and prof. Paul Doukhan for their guidance during my PhD thesis.

Chapter 2

State of the art

In this chapter, firstly we present some of the most commonly used definitions of long memory. Next, we briefly review the (nonlinear) long memory processes studied in the literature. Finally, a short overview on the estimation of parameters in these models is presented at the end of the chapter.

2.1 Long memory

Long memory processes were studied in a literature by numerous authors, see, e.g., the monographs Beran (1997), Doukhan et al. (2003), Giraitis et al. (2012) and the references therein. The problems considered include the detection of long memory, estimation of long memory parameter, limit theorems for long memory processes, simulation of processes, etc. According to Samorodnitsky (2007), the first attempts to study long memory start with the papers of Mandelbrot and his coleagues in 1960s (see Mandelbrot (1965), Mandelbrot and Van Ness (1968)) where the authors seek to explain the phenomenon observed by Hurst (1951) in the empirical data of Nile flows.

Hurst (1951) considered the R/S statistic for Nile river data defined as follows. Given a sequence of n observations $X_1, X_2, \ldots, X_n$, define the partial sum sequence $S_m = X_1 + \cdots + X_m$ for $m = 0, 1, \ldots$ (with $S_0 = 0$). The R/S statistic (range of observations/sample standard deviation) takes the following form

$$\frac{R}{S}(X_1, \ldots, X_n) = \frac{max_{0 \le i \le n}(S_n - \frac{i}{n}S_n) - min_{0 \le i \le n}(S_n - \frac{i}{n}S_n)}{(\frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}S_n)^2)^{1/2}}.$$

It is known that if $X_1, X_2, \ldots$ is a stationary ergodic sequence of random variables with a common mean μ and finite variance, such that the standard central limit theorem holds, then growth rate of R/S statistic is the square root of the sample size. However, Hurst noticed that the growth rate in R/S statistic for the Nile flows data was closer to $n^{0.74}$. This phenomenon was called *Hurst phenomenon* and led to various efforts to explain it.

The fact that the R/S growth rate has unusual behavior suggested that some of the assumptions are not satisfied in the previous example. Mandelbrot (1965) decided to refuse the assumption of the validity of the central limit theorem for process $\{X_t\}$ and proposed to consider a finite variance model with very slowly decaying correlations. This approach appeared to be very succesfull. Fractional Gaussian Noise (the difference of fractional Brownian motion $B_H, H > 0$) defined as

$$X_t := B_H(t) - B_H(t-1),$$

with autocovariances $\mathrm{Cov}(X_{t+n}, X_t) = 2\Big(|n+1|^{2H} + |n-1|^{2H} - 2n^{2H}\Big)$ is the simplest example of such model and gives the growth rate n^H in the R/S statistic.

Since then, various other definitions of the long memory were proposed (see eg. Samorodnitsky (2007), where the most popular definitions are summarized). Most often (due to simplicity and easy estimation from data) the definitions use the second order properties of stochastic processes, for example, asymptotics of covariances, spectral density, variance of partial sums.

The following definition of long memory is based on the slow decay of covariances.

Definition 2.1.1 *A covariance stationary process $\{X(t), t \in Z\}$ is said to have long memory if its autocovariances $\gamma(k) = \mathrm{Cov}(X_t, X_{t-k})$ are not absolutely summable, i.e.*

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| = \infty. \tag{2.1}$$

The process is said to have short memory if

$$\sum_{k=1}^{\infty} |\gamma(k)| < \infty \quad \textit{and} \quad \sum_{k=1}^{\infty} \gamma(k) \neq 0. \tag{2.2}$$

The process is said to have negative memory if

$$\sum_{k=1}^{\infty} |\gamma(k)| < \infty \quad \textit{and} \quad \sum_{k=1}^{\infty} \gamma(k) = 0. \tag{2.3}$$

The definition above imposes very general conditions for autocovariances. It is often useful to go a bit further and specify the decay rate of covariances. Let us first introduce the notion of slowly varying functions.

Definition 2.1.2 *A function L is said to be slowly varying at infinity, if L is positive on $[a;\infty)$, for some $a > 0$, and $\forall t > 0$:*

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1.$$

Definition 2.1.3 *A stationary process $\{X(t), t \in Z\}$ is said to have long memory, if the autocovariance function $\gamma(k) = \mathrm{Cov}(X_t, X_{t-k})$ decays hyperbolically, as $k \to \infty$,*

$$\gamma(k) = k^{2d-1}L(k), \quad 0 < d < 1/2 \tag{2.4}$$

where $L(\cdot)$ is a slowly varying function at infinity. The parameter d is called a long-memory parameter.

Clearly, if the process $\{X(t), t \in \mathbb{Z}\}$ satisfies (2.4), it also satisfies (2.1), the converse not necessarily being truth.

Another common definition of long memory considers the limiting distribution of normalized partial sums process $S_n(\tau) := \sum_{j=1}^{[n\tau]} X_j, \tau \geq 0$ where $[x]$ denotes the integer part of x. Let us first give some definitions (see eg. Giraitis et al. (2016)).

Definition 2.1.4 *(i) A real valued stochastic process $\{Z(t), t \in \mathbb{R}\}$ with $Z(0) = 0$ is said to have stationary increments if for any integer $k > 0$, and for any $t_1 < t_2 < \cdots < t_k$, $t_i \in R$, $i = 1, \ldots, k$ and $h \in \mathbb{R}$, the joint distributions of $\{Z(t_j + h) - Z(h), 1 \leq j \leq k\}$ and $\{Z(t_j) - Z(0), 1 \leq j \leq k\}$ are the same. In other words, if for any $h \in \mathbb{R}$,*

$$\{Z(h + t) - Z(h), t \in \mathbb{R}\} =_{fdd} \{Z(t) - Z(0), t \in \mathbb{R}\}.$$

(ii) A process $\{Z(t), t \in \mathbb{R}\}$ is said to be self-similar with index $H > 0$, if finite dimensional distributions of $\{Z(at)\}$ and $\{a^H Z(t)\}$ are the same for all $a > 0$:

$$\{Z(at), t \in \mathbb{R}\} =_{fdd} \{a^H Z(t), t \in T\}.$$

In other words, for any $a > 0$, $k = 1, 2, \ldots$, and for any $t_j \in \mathbb{R}, 1 \leq j \leq k$,

$$(Z(at_1), \ldots, Z(at_k)) =_{fdd} a^H (Z(t_1), \ldots, Z(t_k)).$$

The process is said to be *H-sssi* if it is self-similar and has stationary increments. A fractional Brownian motion B_H is an example of a Gaussian H-sssi process and plays a very important role in many applications of stochastic processes.

Definition 2.1.5 *Let $0 < H < 1$ be any number. A Gaussian process $B_H = \{B_H(t), t \in \mathbb{R}\}$, with $B_H(0) = 0$, $\mathrm{E}B_H(t) \equiv 0$ and covariance function:*

$$r_H(s,t) := \mathrm{E}B_H(s)B_H(t) = \frac{1}{2}\Big(|s|^{2H} + |t|^{2H} - |s-t|^{2H}\Big), \qquad t, s \geq 0,$$

is called a fractional Brownian motion (fBm) with parameter $0 < H < 1$.

The following result of Lamperti (1962) gives the basis for the definition of distributional long memory.

Theorem 2.1.1 *(Lamperti). Let $\{X_j\}$ be a strictly stationary process and suppose there exist nonrandom numbers $A_n \to \infty$ and $b \in \mathbb{R}$ such that*

$$A_n^{-1} \sum_{j=1}^{[n\tau]} (X_j - b) \to_{fdd.} Z(\tau), \quad \tau \geq 0,$$

where the limit process $Z(\tau), \tau \geq 0$, is not identically zero. Then $\{Z(\tau)\}$ is a stochastically continuous H-sssi process with some parameter $H > 0$ and the normalization $A_n = n^H L(n)$, where $L(\cdot)$ is a slowly varying function .

The process $\{X_t\}$ is said to have *distributional long memory* if the limit process $\{Z(\tau),\ \tau \in [0,1]\}$ has dependent increments. Fractional Brownian motion in Definition 2.1.5 is a typical example of $\{Z(\tau),\ \tau \in [0,1]\}$ in the long memory case.

For more details on various long memory definitions see e.g. Giraitis et al. (2012), Samorodnitsky (2007).

2.2 Long memory processes

The main model for long memory processes is the *linear*, or moving average process

$$X_t = \sum_{s \leq t} b_{t-s}\zeta_s, \qquad t \in \mathbb{Z}, \tag{2.5}$$

where $\{\zeta_s, s \in \mathbb{Z}\}$ is a standardized i.i.d. sequence, and the moving average coefficients b_j decay slowly so that $\sum_{j=0}^{\infty} |b_j| = \infty$, $\sum_{j=0}^{\infty} b_j^2 < \infty$. The last condition guarantees the mean square convergence of series in (2.5) and the process satisfies $\mathrm{E}X_t = 0$, $\mathrm{E}X_t^2 = \sum_{j=0}^{\infty} b_j^2 < \infty$.

In the literature the decay rate of coefficients b_j is often specified. In particular, it is often assumed that

$$b_j \sim \kappa j^{d-1}, \qquad j \to \infty \qquad (\exists\ \kappa > 0,\ 0 < d < 1/2). \tag{2.6}$$

Condition (2.6) guarantees (2.4) and (2.1), i.e. that

$$\gamma(k) \;=\; \sum_{j=0}^{\infty} b_j b_{k+j} \;\sim\; \kappa^2 B(d, 1-2d) k^{-1+2d}, \qquad k \to \infty \tag{2.7}$$

and hence $\sum_{k=1}^{\infty} |\gamma(k)| = \infty$. Thus, $\{X_t\}$ is a long memory process by both Definitions 2.1.3 and 2.1.1 and the parameter d in (1.7) is the long memory parameter of $\{X_t\}$.

A very important case of linear processes (2.5) with (2.6) is the parametric class ARFIMA(p, d, q), in which case $d \in (0, 1/2)$ is the order of fractional integration. The latter class consists of linear processes with coefficients given by power expansion

$$\sum_{j=0}^{\infty} z^j b_j = (1-z)^{-d} \theta(z)/\varphi(z), \qquad |z| < 1$$

where $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q, \varphi(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p$ are polynomials of degree $p, q \geq 0$ that have no common zeros and $\psi(z)$ has no zeros in the unit disc $|z| \leq 1$.

Another important property of the linear process in (2.5)-(2.6) is the distributional long memory or the fact that its (normalized) partial sums process $S_n(\tau) := \sum_{j=1}^{[nt]} X_j$, $\tau \geq 0$ tends to a fractional Brownian motion (Davydov (1970)), viz.,

$$n^{-d-1/2} S_n(\tau) \;\to_{D[0,1]}\; \sigma(d) B_H(t), \tag{2.8}$$

where $H = d + \frac{1}{2}$ is the Hurst parameter, $\sigma(d)^2 := \kappa^2 B(d, 1-2d)/d(1+2d) > 0$ and $\to_{D[0,1]}$ denotes the weak convergence of random processes in the Skorohod space $D[0, 1]$.

Nonlinear long memory processes

Despite its success and popularity, the linear model has its drawbacks as it is not always capable to incorporate the so called “stylized facts” of empirical data, such as clusterings, asymmetry, and various other nonlinearities observed in financial data, together with long memory. As a result, various alternative (nonlinear) long memory models were proposed. We remark that “nonlinear long memory” is a very general term and the literature on this topic is so vast that we will briefly review only some of them.

Subordinated processes

Probably the most studied class of nonlinear long memory processes are subordinated processes of the form $\{Q(X_t)\}$, where $\{X_t\}$ is a stationary Gaussian or linear long memory process and $Q : \mathbb{R} \to \mathbb{R}$ is a nonlinear function (see e.g. Taqqu (1979), Ho and Hsing (1997), Giraitis et al. (2012)). If function $Q : \mathbb{R} \to \mathbb{R}$ is such that $\mathrm{E}Q^2(X_0) < \infty$, then the process $\{Y_t := Q(X_t), t \in \mathbb{Z}\}$ is also stationary. If, in addition, the process $\{X_t\}$ has long memory, we can expect to observe this property in a subordinated process Y_t in a sense of slowly decaying covariances and/or the behavior of partial sums process. However, proving the long memory for subordinated processes is a rather difficult task, mainly because of the nonlinearity of process $\{Y_t, t \in \mathbb{Z}\}$.

The Gaussian process $\{X_t\}$ is probably the only case of subordinated processes for which a complete solution is known. In particular, under specific moment conditions the decay of covariances of process $\{Y_t\}$ is determined by the behavior of process $\{X_t\}$. If $\{X_t\}$ has short memory, a nonlinear function of it also has short memory. However, if $\{X_t\}$ has long memory, the subordinated process $\{Y_t\}$ can have either long or short memory, depending on values of additional parameters. Moreover, central and noncentral limit theorems are already known for this type of processes. The proofs use the method of Hermite expansions, for more details on covariance decay and limit theorems of subordinated Gaussian processes see eg. Giraitis et al. (2012).

Stochastic volatility

A stochastic volatility process $\{r_t\}$ is usually defined as

$$r_t = \sigma_t \zeta_t, \quad t \in \mathbb{Z},$$

where $\{\zeta_t\}$ is a sequence of standartized i.i.d. r.v. and σ_t is a positive function independent of $\{\zeta_t\}$. In contrast to conditionally heteroscedastic models (described in more detail bellow), σ_t is an unobserved process which could be interpreted as volatility but does not represent a conditional variance. The probabilistic properties (stationarity, ergodicity, covariance structure, etc.) of stochastic volatility processes are discussed in a review paper by Davis and Mikosch (2009).

Quite often σ_t is defined as $\sigma_t = f(\eta_t)$ where f is a (nonlinear) function and $\{\eta_t\}$ is some stationary process with well known properties, eg. Gaussian or ARMA (FARIMA) type process. For example, by choosing $f(x) = e^x$ and η_t to be an ARMA(p, q) process we obtain an Exponential GARCH (EGARCH) model proposed by Nelson (1991). When η_t is a FARIMA(p,d,q) process we obtain a Fractional Integrated Exponential GARCH (FIEGARCH) by Bollerslev and Mikkelsen (1996). A related class of long memory stochastic processes was proposed by Breidt et al. (1998) and Harvey (1998) almost simultaneously. This class corresponds to $\eta_t = a + \sum_{j=1}^{\infty} b_j \xi_{t-j}$ where $\xi_t, t \in \mathbb{Z}$ are standard i.i.d r.v. and b_j are the coefficients of FARIMA model. This model is briefly reviewed in Hurvich and Soulier (2009).

Conditional heteroscedasticity

A stationary time series $\{r_t, t \in \mathbb{Z}\}$ is said to be *conditionally heteroscedastic* if its conditional variance $\sigma_t^2 = \mathrm{Var}[r_t | r_s, s < t]$ is a non-constant random process. In financial modeling, r_t are interpreted as (asset) returns and σ_t as volatilities. A class of conditionally heteroscedastic ARCH-type processes is defined from a standardized i.i.d. sequence $\{\zeta_t, t \in \mathbb{Z}\}$ as solutions of stochastic equation

$$r_t = \zeta_t \sigma_t, \qquad \sigma_t = V(r_s, s < t), \tag{2.9}$$

where $V(x_1, x_2, \dots)$ is some function of $x_1, x_2, \dots$.

The ARCH(∞) model corresponds to $V(x_1, x_2, \dots) = \left(a + \sum_{j=1}^{\infty} b_j x_j^2\right)^{1/2}$, or

$$\sigma_t^2 = a + \sum_{j=1}^{\infty} b_j r_{t-j}^2, \tag{2.10}$$

where $a \geq 0, b_j \geq 0$ are coefficients. The ARCH(∞) model includes the well-known ARCH(p) and GARCH(p,q) models of Engle (1982) and Bollerslev (1986). However, despite their tremendous success, the GARCH models are not able to capture some empirical features of asset returns, in particularly, the asymmetric or leverage effect discovered by Black (1976), and the long memory decay in autocorrelation of squares $\{r_t^2\}$. Giraitis and Surgailis (2002) proved that the squared stationary solution of the ARCH(∞) model in (2.10) with $a > 0$ always has short memory, in the sense that $\sum_{j=0}^{\infty} \mathrm{Cov}(r_0^2, r_j^2) < \infty$. (However, for integrated ARCH(∞) models with $\sum_{j=1}^{\infty} b_j = 1, b_j \geq 0$ and $a = 0$ the situation is different; see Giraitis et al. (2016).)

The above shortcomings of the ARCH(∞) model motivated numerous studies proposing alternative forms of the conditional variance and the function $V(x_1, x_2, \dots)$ in (2.9). In particular, stochastic volatility models can display both long memory and leverage except that in their case, the conditional variance is not a function of $r_s, s < t$ alone and therefore it is more difficult to estimate from real data in comparison with the ARCH models; see Shephard and Andersen (2009). Sentana (1995) discussed a class of Quadratic ARCH (QARCH) models with σ_t^2 being a general quadratic form in lagged variables $r_{t-1}, \ldots, r_{t-p}$. Sentana's specification of σ_t^2 encompasses a variety of ARCH models including the asymmetric ARCH model of Engle (1990) and the "linear standard deviation" model of Robinson (1991) corresponding to a case where $a_{ij} = 0$ for $0 < i, j \le p$.

The limiting case (when $p = \infty$) of the "linear standard deviation" (see Robinson (1991)) is the LARCH model discussed in Giraitis et al. (2000) (see also Giraitis and Surgailis (2002), Berkes and Horváth (2003), Giraitis et al. (2004), Giraitis et al. (2009), Truquet (2014) and other papers) and corresponding to $V(x_1, x_2, \ldots) = a + \sum_{j=1}^{\infty} b_j x_j$, or

$$\sigma_t = a + \sum_{j=1}^{\infty} b_j r_{t-j}, \tag{2.11}$$

where $a \in \mathbb{R}, b_j \in \mathbb{R}$ are real-valued coefficients satisfying $B := \left\{ \sum_{j=1}^{\infty} b_j^2 \right\}^{1/2} < \infty$ and $a \neq 0$. Giraitis et al. (2000) showed that a second order strictly stationary solution $\{r_t\}$ to (2.11) exists if and only if $B < 1$, in which case it can be represented by the convergent orthogonal Volterra series

$$r_t = \sigma_t \zeta_t, \quad \sigma_t = a\Big(1 + \sum_{k=1}^{\infty} \sum_{j_1, \ldots, j_k = 1}^{\infty} b_{j_1} \ldots b_{j_k} \zeta_{t-j_1} \ldots \zeta_{t-j_1-\cdots-j_k}\Big). \tag{2.12}$$

Of particular interest is the case when the b_j's in (2.11) are proportional to ARFIMA coefficients, in which case the long memory of the volatility and the (squared) returns can be rigorously proved. In particular, Giraitis et al. (2000) showed that the squared stationary solution $\{r_t^2\}$ of the LARCH model with b_j decaying as in (1.7) under certain moment conditions may have long memory autocorrelations, i.e.

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

where $\kappa_1^2 := \left(\frac{2a\kappa}{1-B^2}\right)^2 B(d, 1-2d) \mathrm{E} r_0^2$. Moreover, its (normalized) partial sums process $S_n(\tau) := \sum_{j=1}^{[nt]} (r_j^2 - \mathrm{E} r_j^2)$, $\tau \ge 0$ tends to a fractional Brownian motion (Giraitis et al. (2000)),

$$n^{-d-1/2} S_n(\tau) \;\to_{D[0,1]}\; \kappa_2 B_H(t), \tag{2.13}$$

where $H = d + \frac{1}{2}$ is the Hurst parameter, $\kappa_2^2 := \kappa_1^2 / (d(1+2d)) > 0$.

The leverage effect in the LARCH model was discussed in detail in Giraitis et al. (2004). Given a stationary conditionally heteroscedastic time series $\{r_t\}$ with $\mathrm{E}|r_t|^3 < \infty$, leverage (a tendency of σ_t^2 to move into the opposite direction as r_s for $s < t$) is usually measured by the covariance $h_{t-s} = \mathrm{Cov}(\sigma_t^2, r_s)$. In Giraitis et al. (2004), the process $\{r_t\}$ is said to have *leverage of order* k ($1 \le k < \infty$) (denoted by $\{r_t\} \in \ell(k)$) whenever

$$h_j < 0, \qquad 1 \le j \le k. \tag{2.14}$$

Given that $\mathrm{E}|r_0|^3 < \infty, |\mu|_3 < \infty$, $B^2 < 1/5$ and $|\mu_3| \le 2(1-5B^2)/B(1+3B^2)$ holds, Giraitis et al. (2004) proved that the second order stationary solution of (2.11) $\{r_t\} \in \ell(k)$ whenever $ab_1 < 0$, $ab_j \le 0$, $j = 2, \ldots, k$, i.o.w. process r_t has leverage of order k.

Despite being able to capture both the asymmetry and the long memory, LARCH model has its drawbacks. The volatility σ_t (2.11) of the LARCH model may assume negative values, lacking some of the usual volatility interpretation and bringing difficulties in parameter estimation.

2.3 Estimation

Let us briefly present the problem of parameter estimation in conditionally heteroscedastic models. Consider a model in (2.9) where σ_t^2 has a parametric form and depends on a parameter $\theta = (\theta_1, \ldots, \theta_k)$. Assume that the observations $\{r_1, r_2, \ldots, r_n\}$, $n \in N$ come from this model with the true parameter $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,k})$. The aim is to get the best possible estimator $\hat{\theta}_n$ of θ_0. The consistency (the fact that $\hat{\theta}_n \to \theta_0$ as $n \to \infty$ in probability) or strong consistency ($\hat{\theta}_n \to \theta_0$ as $n \to \infty$ almost surely) and the asymptotic normality (the convergence of $\hat{\theta}_n - \theta_0$ to Gaussian distribution in law under proper norming) are the desired properties of such estimators.

Let us present Quasi-maximum likelihood (QML) method in more detail as it is the most relevant in this thesis. The idea of QML estimation if to maximize certain objective function that is obtained from the likelihoods of observations under assumption of a particular model. When the "noise" sequence in (2.9) is Gaussian, the maximum likelihood estimator is defined as

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

where Θ is the parameter space and

$$L_n(\theta) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\Big(-\frac{r_t^2}{2\sigma_t^2}\Big)$$

is the likelihood function. The estimator above can be equivalently rewriten as

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n}\Big(\frac{r_t^2}{\sigma_t^2} + \log\sigma_t^2\Big).$$

In the case of non-Gaussian "noise", the same Gaussian likelihood is often used and the estimator is called Quasi-maximum likelihood estimator.

One of the difficulties that arises in the QML estimation is that the volatilities σ_t^2 often depend on infinite past (this is true, for example, for ARCH(∞) in (2.10), LARCH in (2.11), where volatility is written as a linear combination of past returns). However, in practice only a finite number of observations is known. The simplest solution to this problem is to truncate the volatilities by assuming that the unknown returns $\{r_i, i < 0\}$ are all equal to 0. In this case two estimators are often considered: one involving exact conditional variance σ_t^2 depending on infinite past

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} L_n(\theta), \qquad L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\Big(\frac{r_t^2}{\sigma_t^2(\theta)} + \log\sigma_t^2(\theta)\Big)$$

and its more realistic version obtained by replacing σ_t^2 by $\tilde{\sigma}_t^2$ depending only on finite past $(r_s, 1 \le s < t)$:

$$\tilde{\theta}_n = \arg\max_{\theta\in\Theta} \tilde{L}_n(\theta), \qquad \tilde{L}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\Big(\frac{r_t^2}{\tilde{\sigma}_t^2(\theta)} + \log\tilde{\sigma}_t^2(\theta)\Big).$$

Quasi-maximum likelihood method gives consistent and asymptotically normal estimators of parameters in strictly stationary GQARCH model under very mild regularity conditions and does not require any conditions for higher moments (see Francq and Zakoian (2009), Chapter 7, Theorems 7.1 and 7.2). The latter fact is particularly relevant from the practical perspective as the requirements of finite fourth or even higher moments seem to be too strict in real data, for example in financial time series. Robinson and Zaffaroni (2006) proved strong consistency and asymptotic normality of QML estimator in ARCH(∞) model (2.10) with $a > 0$ and the coefficients $b_j = b_j(\lambda)$ written as some functions depending on finite dimensional parameter $\lambda \in \mathbb{R}^m$. The estimation of parameters in long memory LARCH model was studied in Beran et al. (2013). Recall that the main dissadvantage of LARCH model is the fact that the volatilities might become negative and in general are not separated from 0. Thus, the standard Quasi-maximum likelihood estimator is inconsistent. Beran and Schützner (2009) considered a modified Quasi-maximum likelihood estimator that involves an additional "small" tuning parameter ϵ, also other estimation methods for LARCH model were developed in Francq and Zakoian (2010b), Levine et al. (2009), Truquet (2014).

Finally, many other methods for estimation were discussed in the literature, for example in Straumann (2005), Francq and Zakoian (2009) and we will only mention some of them. Probably the simplest method for ARCH models is the Least squares (LS) method that is based on the minimization of squared errors. LS provides the estimators in ARCH case explicitely, moreover, they are consistent and asymptotically normal if $\mathrm{E}r_t^4 < \infty$ and $\mathrm{E}r_t^8 < \infty$ respectively (see Francq and Zakoian (2009) for more details). Whittle (1953) proposed estimator based on spectral densities and periodograms. It is often used in practice and also covers long memory cases. We are not going into more details here as in this thesis we are mainly focusing on QML estimation.

Chapter 3

Projective stochastic equations and nonlinear long memory

Abstract. A projective moving average $\{X_t, t \in \mathbb{Z}\}$ is a Bernoulli shift written as a backward martingale transform of the innovation sequence. We introduce a new class of nonlinear stochastic equations for projective moving averages, termed projective equations, involving a (nonlinear) kernel Q and a linear combination of projections of X_t on "intermediate" lagged innovation subspaces with given coefficients $\alpha_i, \beta_{i,j}$. The class of such equations include usual moving-average processes and the Volterra series of the LARCH model. Solvability of projective equations is obtained using a recursive equality for projections of the solution X_t. We show that under certain conditions on $Q, \alpha_i, \beta_{i,j}$, this solution exhibits covariance and distributional long memory, with fractional Brownian motion as the limit of the corresponding partial sums process.

3.1 Introduction

The present chapter introduces a new class of nonlinear processes which generalize the linear model in (2.5)-(2.6) and enjoy similar long memory properties to (2.7) and (2.8). These processes are defined through solutions of the so-called *projective stochastic equations.* Here, the term "projective" refers to the fact that these equations contain linear combinations of projections, or conditional expectations, of X_t's on lagged innovation subspaces which enter the equation in a nonlinear way.

Let us explain the main idea of our construction. We call a *projective moving average* a random process $\{X_t\}$ of the form

$$X_t \;=\; \sum_{s \le t} g_{s,t} \zeta_s, \qquad t \in \mathbb{Z}, \tag{3.1}$$

where $\{\zeta_s\}$ is a sequence of standardized i.i.d. r.v.'s as in (1.6), $g_{t,t} \equiv g_0$ is a deterministic constant and $g_{s,t}$, $s < t$ are r.v.'s depending only on $\zeta_{s+1}, \ldots, \zeta_t$ such that

$$g_{s,t} = g_{t-s}(\zeta_{s+1}, \ldots, \zeta_t), \qquad s < t, \tag{3.2}$$

where $g_j : \mathbb{R}^j \to \mathbb{R}$, $j = 1, 2, \ldots$ are *nonrandom* functions satisfying

$$\sum_{s \le t} \mathrm{E} g_{s,t}^2 \;=\; \sum_{s \le 0} \mathrm{E} g_{-s}^2(\zeta_{s+1}, \ldots, \zeta_0) \;<\; \infty. \tag{3.3}$$

It follows easily that under condition (3.3) the series in (3.1) converges in mean square and define a stationary process with zero mean and finite variance $\mathrm{E}X_t^2 = \sum_{s \le t} \mathrm{E} g_{s,t}^2$. The next question - how to choose the "coefficients" $g_{s,t}$ (3.2) so that they depend on X_t and behave like (2.6) when $j = t - s \to \infty$?

A particularly simple choice of the $g_{s,t}$'s to achieve the above goals is

$$g_{s,t} \;=\; b_{t-s} Q(\mathrm{E}_{[s+1,t]} X_t), \qquad s \le t \tag{3.4}$$

where b_j are as in (2.6), $Q : \mathbb{R} \to \mathbb{R}$ is a given deterministic kernel, and $\mathrm{E}_{[s+1,t]} X_t := \mathrm{E}[X_t | \zeta_v, s+1 \le v \le t]$ is the projection of X_t onto the subspace of L^2 generated by the innovations $\zeta_v, s+1 \le v \le t$ (the conditional expectation). The corresponding projective stochastic equation has the form

$$X_t \;=\; \sum_{s \le t} b_{t-s} Q(\mathrm{E}_{[s+1,t]} X_t) \zeta_s. \tag{3.5}$$

Notice that when $s \to -\infty$ then $\mathrm{E}_{[s+1,t]} X_t \to X_t$ by a general property of a conditional expectation and then $g_{s,t} \sim b_{t-s} Q(X_t)$ if Q is continuous. This means that the $g_{s,t}$'s in (3.4) feature both the long memory in (2.6) and the dependence on the "current" value X_t through $Q(X_t)$. In particular, for $Q(x) = \max(0, x)$, the behavior of $g_{s,t}$ in (3.4) strongly depends on the sign of X_t and the trajectory of (3.5) appears to be very asymmetric (see Figure 3.3, top).

Let us briefly describe the remaining sections. Section 3.2 contains basic definitions and properties of projective processes. Section 3.3 introduces a general class of projective stochastic equations, (3.5) being a particular case. We obtain sufficient conditions of solvability of these equations, and a recurrent formula for computation of "coefficients" $g_{s,t}$ (Theorem 3.3.1). Sections 3.4 and 3.5 present some examples and simulated trajectories and histograms of projective equations. It turns out that the LARCH model studied in Giraitis et al. (2000) and elsewhere is a particular case of projective equations corresponding to linear kernel $Q(x)$ (Section 3.4). Some modifications of projective equations are discussed in Section 3.6. Section 3.7 deals with long memory properties of stationary solutions of stochastic projective equations. We show that under some additional conditions these solutions have long memory properties similar to (2.7) and (2.8).

Finally, we remark that "nonlinear long memory" is a general term and that other time series models different from ours for such behavior were proposed in the literature. Among them, probably the most studied class are subordinated processes of the form $\{Q(X_t)\}$, where $\{X_t\}$ is a Gaussian or linear long memory process and $Q : \mathbb{R} \to \mathbb{R}$ is a nonlinear function. See Taqqu (1979), Ho and Hsing (1997) and Giraitis et al. (2012) for a detailed discussion. A related class of Gaussian subordinated stochastic volatility models is studied in Robinson (2001). Doukhan et al. (2012) discuss a class of long memory Bernoulli shifts. Baillie and Kapetanios (2008) consider fractionally integrated process with nonlinear autoregressive innovations. A general invariance principle for fractionally integrated models with weakly dependent innovations satisfying the projective dependence condition of Wu (2005) is established in Shao and Wu (2006). See also Wu and Min (2005) and Remark 3.7.1 below.

We expect that the results of this chapter can be extended in several directions, e.g., projective equations with initial condition, continuous time processes, random field set-up, infinite variance processes. For applications, a major challenge is estimation of parameters of projective equations.

3.2 Projective processes and their properties

Let $\{\zeta_t, t \in \mathbb{Z}\}$ be a sequence of i.i.d. r.v.'s with $\mathrm{E}\zeta_0 = 0$, $\mathrm{E}\zeta_0^2 = 1$. For any integers $s \le t$ we denote $\mathcal{F}_{[s,t]} := \sigma\{\zeta_u : u \in [s,t]\}$ the sigma-algebra generated by $\zeta_u, u \in [s,t]$, $\mathcal{F}_{(-\infty,t]} := \sigma\{\zeta_u : u \le t\}$, $\mathcal{F} := \sigma\{\zeta_u : u \in \mathbb{Z}\}$. For $s > t$, we define $\mathcal{F}_{[s,t]} := \{\emptyset, \Omega\}$ as the trivial sigma-algebra. Let $L^2_{[s,t]}$, $L^2_{(-\infty,t]}$, L^2 be the spaces of all square integrable r.v.'s ξ measurable w.r.t. $\mathcal{F}_{[s,t]}$, $\mathcal{F}_{(-\infty,t]}$, $\mathcal{F}$, respectively. For any $s, t \in \mathbb{Z}$ let

$$\mathrm{E}_{[s,t]}[\xi] := \mathrm{E}\left[\xi \middle| \mathcal{F}_{[s,t]}\right], \qquad \xi \in L^2$$

be the conditional expectation. Then $\xi \mapsto \mathrm{E}_{[s,t]}[\xi]$ is a bounded linear operator in L^2; moreover, $\mathrm{E}_{[s,t]}$, $s, t \in \mathbb{Z}$ is a projection family satisfying $\mathrm{E}_{[s_2,t_2]}\mathrm{E}_{[s_1,t_1]} = \mathrm{E}_{[s_2,t_2]\cap[s_1,t_1]}$ for any intervals $[s_1, t_1], [s_2, t_2] \subset \mathbb{Z}$. From the definition of conditional expectation it follows that if $g_u : \mathbb{R} \to \mathbb{R}, u \in \mathbb{Z}$ are arbitrary measurable functions with $\mathrm{E}g_u^2(\zeta_u) < \infty$, $[s_2, t_2] \subset \mathbb{Z}$ is a given interval and $\xi = \prod_{u\in[s_2,t_2]} g_u(\zeta_u)$ is a product of independent r.v.'s, then for any interval $[s_1, t_1] \subset \mathbb{Z}$

$$\mathrm{E}_{[s_1,t_1]} \prod_{u \in [s_2,t_2]} g_u(\zeta_u) \;=\; \prod_{u \in [s_1,t_1] \cap [s_2,t_2]} g_u(\zeta_u) \prod_{v \in [s_2,t_2] \setminus [s_1,t_1]} \mathrm{E}[g_v(\zeta_v)].$$

In particular, if $\mathrm{E} g_u(\zeta_u) = 0$, $u \in \mathbb{Z}$ then

$$\mathrm{E}_{[s_1,t_1]} \prod_{u \in [s_2,t_2]} g_u(\zeta_u) \;=\; \begin{cases} \prod_{u \in [s_1,t_1]} g_u(\zeta_u), & [s_2,t_2] \subset [s_1,t_1], \\ 0, & [s_2,t_2] \not\subset [s_1,t_1]. \end{cases} \tag{3.6}$$

Any r.v. $Y_t \in L^2_{(-\infty,t]}$ can be expanded into orthogonal series $Y_t = \mathrm{E}Y_t + \sum_{s \le t} P_{s,t} Y_t$, where $P_{s,t} Y_t := (\mathrm{E}_{[s,t]} - \mathrm{E}_{[s+1,t]}) Y_t$. Note that $\{P_{s,t} Y_t, \mathcal{F}_{s,t}, s \le t\}$ is a backward martingale difference sequence and $\mathrm{E}Y_t^2 = (\mathrm{E}Y_t)^2 + \sum_{s \le t} \mathrm{E}(P_{s,t} Y_t)^2$.

Definition 3.2.1 *A projective process is a random sequence $\{Y_t \in L^2_{(-\infty,t]},\ t \in \mathbb{Z}\}$ of the form*

$$Y_t \;=\; \mathrm{E}Y_t + \sum_{s \le t} g_{s,t} \zeta_s, \tag{3.7}$$

where $g_{s,t}$ are r.v.'s satisfying the following conditions (i) and (ii):

(i) $g_{s,t}$ is $\mathcal{F}_{[s+1,t]}$-measurable, $\forall s, t \in \mathbb{Z}$, $s < t$; $g_{t,t}$ is a deterministic number;

(ii) $\sum_{s \le t} \mathrm{E} g^2_{s,t} < \infty$, $\forall\, t \in \mathbb{Z}$.

In other words, a projective process has the property that the projections $\mathrm{E}_{[s,t]} Y_t = \mathrm{E}Y_t + \sum_{i=s}^t P_{i,t} Y_t = \mathrm{E}Y_t + \sum_{i=s}^t \zeta_i g_{i,t}$, $s \le t$ form a backward martingale transform w.r.t. the nondecreasing family $\{\mathcal{F}_{[s,t]}, s \le t\}$ of sigma-algebras, for each $t \in \mathbb{Z}$ fixed. A consequence of the last fact is the following moment inequality which is an easy consequence of Rosenthal's inequality (Hall and Heyde (1980), p.24). See also Giraitis et al. (2012), Lemma 2.5.2.

Proposition 3.2.1 *Let $\{Y_t\}$ be a projective process in (3.7). Assume that $\mu_p := \mathrm{E}|\zeta_0|^p < \infty$ and $\sum_{s \le t} (\mathrm{E}|g_{s,t}|^p)^{2/p} < \infty$ for some $p \ge 2$. Then $\mathrm{E}|Y_t|^p < \infty$. Moreover, there exists a constant $C_p < \infty$ depending on p alone and such that*

$$\mathrm{E}|Y_t|^p \;\le\; C_p \Big(|\mathrm{E}Y_t|^p + \mu_p \Big(\sum_{s \le t} (\mathrm{E}|g_{s,t}|^p)^{2/p} \Big)^{p/2} \Big).$$

Definition 3.2.2 *A projective moving average is a projective process of (3.7) such that the mean $\mathrm{E}Y_t = \mu$ is constant and there exist a number $g_0 \in \mathbb{R}$ and nonrandom measurable functions $g_j : \mathbb{R}^j \to \mathbb{R}, j = 1, 2, \ldots$ such that*

$$g_{s,t} = g_{t-s}(\zeta_{s+1}, \ldots, \zeta_t) \qquad \textit{a.s., for any} \quad s \le t,\ s, t \in \mathbb{Z}.$$

By definition, a projective moving average is a stationary Bernoulli shift (Dedecker et al. (2007), p.21):

$$Y_t \;=\; \mu + \sum_{s\le t} \zeta_s g_{t-s}(\zeta_{s+1},\ldots,\zeta_t) \tag{3.8}$$

with mean μ and covariance

$$\begin{aligned}\mathrm{Cov}(Y_s, Y_t) &= \sum_{u\le s} \mathrm{E}[g_{s-u}(\zeta_{u+1},\ldots,\zeta_s) g_{t-u}(\zeta_{u+1},\ldots,\zeta_t)] \\ &= \sum_{u\le 0} \mathrm{E}[g_{-u}(\zeta_{u+1},\ldots,\zeta_0) g_{t-s-u}(\zeta_{u+1},\ldots,\zeta_{t-s-u})], \quad s \le t. \end{aligned} \tag{3.9}$$

These facts together with the ergodicity of Bernoulli shifts (implied by a general result in Stout (1974), Theorem 3.5.8) are summarized in the following corollary.

Corollary 3.2.1 *A projective moving average is a strictly stationary and ergodic stationary process with finite variance and covariance given in (3.9).*

Remark 3.2.1 If the coefficients $g_{s,t}$ are nonrandom, a projective moving average is a linear process $Y_t = \mu + \sum_{s\le t} g_{t-s}\zeta_s$, $t \in \mathbb{Z}$.

Proposition 3.2.2 *Let $\{Y_t\}$ be a projective process of (3.7) and $\{a_j, j \ge 0\}$ a deterministic sequence, $\sum_{j=0}^{\infty} |a_j| < \infty$, $\sum_{j=0}^{\infty} |a_j||\mathrm{E}Y_{t-j}| < \infty$. Then $\{u_t := \sum_{j=0}^{\infty} a_j Y_{t-j}$, $t \in \mathbb{Z}\}$ is a projective process $u_t = \mathrm{E}u_t + \sum_{s\le t} \zeta_s G_{s,t}$ with $\mathrm{E}u_t = \sum_{j=0}^{\infty} a_j \mathrm{E}Y_{t-j}$ and coefficients $G_{s,t} := \sum_{j=0}^{t-s} a_j g_{s,t-j}$.*

Proof follows easily by the Cauchy-Schwarz inequality and is omitted. □

Proposition 3.2.3 *If $\{Y_t\}$ is a projective process of (3.7), then for any $s \le t$*

$$\mathrm{E}_{[s,t]} Y_t \;=\; \mathrm{E}Y_t + \sum_{s\le u\le t} \zeta_u g_{u,t}, \quad P_{s,t} Y_t \;=\; (\mathrm{E}_{[s,t]} - \mathrm{E}_{[s+1,t]}) Y_t \;=\; \zeta_s g_{s,t}. \tag{3.10}$$

The representation (3.7) is unique: if (3.7) and $Y_t = \sum_{s\le t} g'_{s,t}\zeta_s$ are two representations, with $g'_{s,t}$ satisfying conditions (i) and (ii) of Definition 3.2.1, then $g'_{s,t} = g_{s,t}\ \forall\ s \le t$.

Proof of (3.10) is immediate by definition of projective process. From (3.10) it follows that $\zeta_s g''_{s,t} = 0$, where $g''_{s,t} := g_{s,t} - g'_{s,t}$ is independent of ζ_s. Relation $\mathrm{E}\zeta_s^2 = 1$ implies $\mathrm{P}(|\zeta_s|^2 > \epsilon) > 0$ for all $\epsilon > 0$ small enough. Hence, $0 = \mathrm{P}(|\zeta_s g''_{s,t}| > \epsilon) \ge \mathrm{P}(|\zeta_s| > \sqrt{\epsilon}, |g''_{s,t}| > \sqrt{\epsilon}) = \mathrm{P}(|\zeta_s| > \sqrt{\epsilon})\mathrm{P}(|g''_{s,t}| > \sqrt{\epsilon})$, implying $\mathrm{P}(|g''_{s,t}| > \sqrt{\epsilon}) = 0$ for any $\epsilon > 0$. □

The following invariance principle is due to Dedecker and Merlevède (Dedecker and Merlevède (2003), Cor. 3), see also (Wu (2005), Theorem 3 (i)).

Proposition 3.2.4 *Let $\{Y_t\}$ be a projective moving average of (3.7) such that $\mu = 0$ and*

$$\Omega(2) \;:=\; \sum_{t=0}^{\infty} \|g_{0,t}\| \;<\infty, \tag{3.11}$$

where $\|\xi\| = \mathrm{E}^{1/2}[\xi^2]$, $\xi \in L^2$. Then

$$n^{-1/2}\sum_{t=1}^{[n\tau]} Y_t \;\longrightarrow_{D[0,1]}\; c_Y B(\tau), \tag{3.12}$$

where B is a standard Brownian motion and $c_Y^2 := \|\sum_{t=0}^{\infty} g_{0,t}\|^2 = \sum_{t\in\mathbb{Z}} \mathrm{E}[Y_0 Y_t]$.

3.3 Projective stochastic equations

Let $Q_{s,t} = Q_{s,t}(x_{u,v}, s < u \le v \le t)$, $s, t \in \mathbb{Z}, s < t$ be some given measurable deterministic functions depending on $(t-s)(t-s+1)/2$ real variables $x_{u,v}$, $s < t$, and μ_t, $Q_{t,t}$, $t \in \mathbb{Z}$ be some given constants. A *projective stochastic equation* has the form

$$X_t \;=\; \mu_t + \sum_{s\le t} \zeta_s Q_{s,t}(\mathrm{E}_{[u,v]}X_v, s < u \le v \le t). \tag{3.13}$$

Definition 3.3.1 *By solution of (3.13) we mean a projective process $\{X_t, t \in \mathbb{Z}\}$ satisfying $\sum_{s\le t} \mathrm{E}[Q^2_{s,t}(\mathrm{E}_{[u,v]}X_v, s < u \le v \le t)] \;<\; \infty$ and (3.13) for any $t \in \mathbb{Z}$.*

Proposition 3.3.1 *Assume that that $\mu_t = \mu$ does not depend on $t \in \mathbb{R}$, the functions $Q_{s,t} = Q_{t-s}$, $s \le t$ in (3.13) depend only on $t - s$, and that $\{X_t\}$ is a solution of (3.13). Then $\{X_t\}$ is a projective moving average of (3.8) with $\mathrm{E}X_t = \mu$ and $g_n : \mathbb{R}^n \to \mathbb{R}$, $n = 0, 1, \ldots$ defined recursively by*

$$g_0 \;:=\; Q_0, \tag{3.14}$$
$$g_n(x_{-n+1}, \ldots, x_0) \;:=\; Q_n\Big(\mu + \sum_{k=u}^{v} x_k g_{v-k}(x_{u+1}, \ldots, x_v), -n < u \le v \le 0\Big), \quad n \ge 1.$$

Moreover, such solution is unique.

Proof. From (3.13) and the uniqueness of (3.7) (Proposition 3.2.3) we have $X_t = \mu + \sum_{s\le t} g_{s,t}\zeta_s$, where $g_{s,t} = Q_{t-s}(\mathrm{E}_{[u,v]}X_v, s < u \le v \le t)$. For $s = t$ this yields $g_{t,t} = Q_0 = g_0\, \forall t \in \mathbb{Z}$ as in (3.14). Similarly, $g_{t-1,t} = Q_1(\mathrm{E}_{[t,t]}X_t) = Q_1(\mu + g_0\zeta_t) = g_1(\zeta_t)$, where g_1 is defined in (3.14). Assume by induction that

$$g_{t-m,t} = g_m(\zeta_{t-m+1}, \ldots, \zeta_t), \qquad \forall\, t \in \mathbb{Z} \tag{3.15}$$

with g_m defined in (3.14), hold for any $0 \le m < n$ and some $n \ge 1$; we need to show that (3.15) holds for $m = n$, too. Using (3.15), (3.10) and (3.14) we obtain

$$\begin{aligned} g_{t-n,t} &= Q_n(\mathrm{E}_{[u,v]} X_v, t-n < u \le v \le t) \\ &= Q_n\Big(\mu + \sum_{k=u}^{v} \zeta_k g_{v-k}(\zeta_{u+1}, \ldots, \zeta_v), t-n < u \le v \le t\Big) \\ &= g_n(\zeta_{t-n+1}, \ldots, \zeta_t). \end{aligned}$$

This proves the induction step $n-1 \to n$ and hence the proposition, too, since the uniqueness follows trivially. $\square$

Clearly, the choice of possible kernels $Q_{s,t}$ in (3.13) is very large. In this chapter we focus on the following class of projective stochastic equations:

$$X_t = \mu + \sum_{s \le t} \zeta_s Q\Big(\alpha_{t-s} + \sum_{s<u\le t} \beta_{t-u,u-s}\left(\mathrm{E}_{[u,t]} X_t - \mathrm{E}_{[u+1,t]} X_t\right)\Big), \qquad (3.16)$$

where $\{\alpha_i, i \ge 0\}$, $\{\beta_{i,j}, i \ge 0, j \ge 1\}$ are given arrays of real numbers, $\mu \in \mathbb{R}$ is a constant, and $Q = Q(x)$ is a measurable function of a single variable $x \in \mathbb{R}$. Two modifications of (3.16) are briefly discussed below, see (3.38) and (3.41). Particular cases of (3.16) are

$$X_t = \sum_{s \le t} \zeta_s Q\Big(\alpha_{t-s} + \beta_{t-s} \mathrm{E}_{[s+1,t]} X_t\Big), \qquad (3.17)$$

and

$$X_t = \mu + \sum_{s \le t} \zeta_s Q\Big(\alpha_{t-s} + \sum_{s<u\le t} \beta_{u-s}\left(\mathrm{E}_{[u,t]} X_t - \mathrm{E}_{[u+1,t]} X_t\right)\Big), \qquad (3.18)$$

corresponding to $\beta_{i,j} = \beta_{i+j}$ and $\beta_{i,j} = \beta_j$, respectively.

Next, we study the solvability of projective equation (3.16). We assume that Q satisfies the following dominating bound: there exists a constant $c_Q > 0$ such that

$$|Q(x)| \le c_Q |x|, \qquad \forall\, x \in \mathbb{R}. \qquad (3.19)$$

Denote

$$K_Q := \sum_{i=0}^{\infty} \alpha_i^2 \sum_{k=0}^{\infty} c_Q^{2k+2} \sum_{j_1=1}^{\infty} \beta_{i,j_1}^2 \cdots \sum_{j_k=1}^{\infty} \beta_{i+j_1+\cdots+j_{k-1},j_k}^2. \qquad (3.20)$$

The main result of this section is the following theorem:

Theorem 3.3.1 *(i) Assume condition (3.19) and*

$$K_Q \ < \ \infty. \tag{3.21}$$

Then there exists a unique solution $\{X_t\}$ of (3.16), which is written as a projective moving average in (3.7) with coefficients $g_{t-k,t}$ recursively defined as

$$g_{t-k,t} \ := \ \begin{cases} Q\Big(\alpha_k + \sum_{i=0}^{k-1} \beta_{i,k-i}\zeta_{t-i} g_{t-i,t}\Big), & k = 1, 2, \ldots, \\ Q(\alpha_k), & k = 0. \end{cases} \tag{3.22}$$

More explicitly,

$$\begin{aligned} X_t \ &= \ \mu + Q(\alpha_0)\zeta_t + Q\Big(\alpha_1 + \beta_{0,1}\zeta_t Q(\alpha_0)\Big)\zeta_{t-1} \\ &+ \ Q\Big(\alpha_2 + \beta_{0,2}\zeta_t Q(\alpha_0) + \beta_{1,1}\zeta_{t-1} Q\Big(\alpha_1 + \beta_{0,1}\zeta_t Q(\alpha_0)\Big)\Big)\zeta_{t-2} + \ldots . \end{aligned}$$

(ii) In the case of linear function $Q(x) = c_Q x$, condition (3.21) is also necessary for the existence of a solution of (3.16).

Proof. (i) Let us show that the $g_{k-t,t}$'s as defined in (3.22) satisfy $\sum_{k=0}^{\infty} \mathrm{E} g_{t-k,t}^2 < \infty$. From (3.19) and (3.22) we have the recurrent inequality:

$$\mathrm{E} g_{t-k,t}^2 \ \le \ c_Q^2 \mathrm{E}\Big(\alpha_k + \sum_{i=0}^{k-1} \beta_{i,k-i}\zeta_{t-i} g_{t-i,t}\Big)^2 \ = \ c_Q^2\Big(\alpha_k^2 + \sum_{i=0}^{k-1} \beta_{i,k-i}^2 \mathrm{E} g_{t-i,t}^2\Big). \tag{3.23}$$

Iterating (3.23) we obtain

$$\begin{aligned} \mathrm{E} g_{t-k,t}^2 \ &\le \ c_Q^2\Big(\alpha_k^2 + c_Q^2 \sum_{i=0}^{k-1} \beta_{i,k-i}^2\Big(\alpha_i^2 + \sum_{j=0}^{i-1} \beta_{j,i-j}^2 \mathrm{E} g_{t-j,t}^2\Big)\Big) \\ &= \ c_Q^2 \alpha_k^2 + c_Q^4 \sum_{i=0}^{k-1} \alpha_i^2 \beta_{i,k-i}^2 + c_Q^6 \sum_{i=0}^{k-1} \alpha_i^2 \sum_{j_1=1}^{k-1-i} \beta_{i,j_1}^2 \beta_{i+j_1,k-i-j_1}^2 + \ldots \end{aligned} \tag{3.24}$$

and hence

$$\begin{aligned} \sum_{k=0}^{\infty} \mathrm{E} g_{t-k,t}^2 \ &\le \ c_Q^2 \sum_{i=0}^{\infty} \alpha_i^2 + c_Q^4 \sum_{i=0}^{\infty} \alpha_i^2 \sum_{j_1=1}^{\infty} \beta_{i,j_1}^2 + c_Q^6 \sum_{i=0}^{\infty} \alpha_i^2 \sum_{j_1=1}^{\infty} \beta_{i,j_1}^2 \sum_{j_2=1}^{\infty} \beta_{i+j_1,j_2}^2 + \ldots \\ &= \ K_Q \ < \ \infty \end{aligned} \tag{3.25}$$

according to (3.21). Therefore, $X_t = \mu + \sum_{s \le t} g_{s,t}\zeta_s$ is a well-defined projective moving-average. The remaining statements about X_t follow from Proposition 3.3.1.

(ii) Similarly to (3.23), (3.25) in the case $Q(x) = c_Q x$ we obtain

$$\mathrm{E} g_{t-k,t}^2 \;=\; c_Q^2 \mathrm{E}\Big(\alpha_k + \sum_{i=0}^{k-1} \beta_{i,k-i}\zeta_{t-i} g_{t-i,t}\Big)^2 \;=\; c_Q^2\Big(\alpha_k^2 + \sum_{i=0}^{k-1} \beta_{i,k-i}^2 \mathrm{E} g_{t-i,t}^2\Big)$$

and hence $\mathrm{Var}(X_t) = \sum_{k=0}^\infty \mathrm{E} g_{t-k,t}^2 = K_Q$. This proves (ii) and Theorem 3.3.1, too. $\square$

Remark 3.3.1 From recurrent relation (3.22), the $g_{t-k,t}$'s can be expressed as functions of $\zeta_{t-k+1}, \dots, \zeta_t$ via the so-called nested Volterra series (see Appendix B and the extented version of Grublytė and Surgailis (2014) available at arXiv:1312.1938).

In the case of equations (3.17) and (3.18), condition (3.21) can be simplified, see below. Note that for $A^2 := \sum_{i=0}^\infty \alpha_i^2 = 0$, equations (3.22) admit a trivial solution $g_{t-k,t} = 0$ since $Q(0) = 0$ by (3.19), leading to the constant process $X = \mu$ in (3.16).

Proposition 3.3.2 *(i) Let* $A^2 > 0$, $\beta_{i,j} = \beta_{i+j}$, $i \ge 0$, $j \ge 1$, *and* $B^2 := \sum_{i=0}^\infty \beta_i^2$. *Then* $K_Q < \infty$ *is equivalent to* $A^2 < \infty$ *and* $B^2 < \infty$.
(ii) Let $A^2 > 0$, $\beta_{i,j} = \beta_j$, $i \ge 0$, $j \ge 1$ *and* $B^2 := \sum_{i=1}^\infty \beta_i^2$. *Then* $K_Q < \infty$ *is equivalent to* $A^2 < \infty$ *and* $c_Q^2 B^2 < 1$. *Moreover,* $K_Q = c_Q^2 A^2/(1 - c_Q^2 B^2)$.

Proof. (i) By definition,

$$\begin{aligned} K_Q &= \sum_{k=0}^\infty c_Q^{2k+2} \sum_{i=0}^\infty \alpha_i^2 \sum_{j_1=1}^\infty \beta_{i+j_1}^2 \cdots \sum_{j_k=1}^\infty \beta_{i+j_1+\cdots+j_{k-1}+j_k}^2 \\ &= \sum_{k=0}^\infty c_Q^{2k+2} \sum_{0 \le i < j_1 < \cdots < j_k < \infty} \alpha_i^2 \beta_{j_1}^2 \dots \beta_{j_k}^2 \\ &\le \sum_{k=0}^\infty c_Q^{2k+2} A^2 B_1^2 \dots B_k^2, \end{aligned}$$

where $B_k^2 := \sum_{j=k}^\infty \beta_j^2$. Since $B^2 < \infty$ entails $\lim_{k\to\infty} B_k^2 = 0$, $\forall \epsilon > 0 \,\exists\, K \ge 1$ such that $B_k^2 < \epsilon / c_Q^2 \;\forall\, k > K$. Hence,

$$K_Q \;\le\; c_Q^2 A^2 \Big(\sum_{k=0}^K (c_Q^2 B^2)^k + \sum_{k=K}^\infty \epsilon^k \Big) \;<\infty.$$

Therefore, $A^2 < \infty$ and $B^2 < \infty$ imply $K_Q < \infty$. The converse implication is obvious.

(ii) Follows by

$$K_Q = \sum_{k=0}^{\infty} c_Q^{2k+2} \sum_{i=0}^{\infty} \alpha_i^2 \sum_{j_1=1}^{\infty} \beta_{j_1}^2 \cdots \sum_{j_k=1}^{\infty} \beta_{j_k}^2 = \sum_{k=0}^{\infty} c_Q^{2k+2} A^2 (B^2)^k = \frac{c_Q^2 A^2}{1 - c_Q^2 B^2}.$$

$\square$

Remark 3.3.2 It is not difficult to show that conditions on the $\beta_{i,j}$'s in Proposition 3.3.2 (i) and (ii) are part of the following more general condition:

$$\limsup_{i\to\infty} \sum_{j=1}^{\infty} c_Q^2 \beta_{i,j}^2 < 1,$$

which also guarantees that $K_Q < \infty$.

The following Proposition 3.3.3 obtains a sufficient condition for the existence of higher moments $\mathrm{E}|X_t|^p < \infty, p > 2$ of the solution of projective equation (3.16). The proof of Proposition 3.3.3 is based on a recurrent use of Rosenthal-type inequality of Proposition 3.2.1, which contains an absolute constant C_p depending only on p. For $p \geq 2$, denote

$$K_{Q,p} := C_p^{2/p} \sum_{i=0}^{\infty} \alpha_i^2 \sum_{k=0}^{\infty} (c_Q C_p^{1/p} \mu_p^{1/p})^{2k+2} \sum_{j_1=1}^{\infty} \beta_{i,j_1}^2 \cdots \sum_{j_k=1}^{\infty} \beta_{i+j_1+\cdots+j_{k-1},j_k}^2, \tag{3.26}$$

where (recall) $\mu_p = \mathrm{E}|\zeta_0|^p$. Note $C_2 = \mu_2 = 1$, hence $K_{Q,2} = K_Q$ coincides with (3.20).

Proposition 3.3.3 *Assume conditions of Theorem 3.3.1 and $K_{Q,p} < \infty$, for some $p \geq 2$. Then $\mathrm{E}|X_t|^p < \infty$.*

Proof. The proof is similar to that of Theorem 3.3.1 (i). By Proposition 3.2.1,

$$\begin{aligned} \Big(\mathrm{E}\Big|X_t\Big|^p\Big)^{2/p} &\leq C_p^{2/p}\Big(\Big|\mathrm{E}X_t\Big|^p + \mu_p\Big(\sum_{s\leq t}\Big(\mathrm{E}|g_{s,t}|^p\Big)^{2/p}\Big)^{p/2}\Big)^{2/p} \\ &= C_p^{2/p}\mu_p^{2/p}\sum_{s\leq t}(\mathrm{E}|g_{s,t}|^p)^{2/p}. \end{aligned}$$

Using condition (3.19), Proposition 3.2.1 and inequality $(a+b)^q \leq a^q + b^q$, $0 < q \leq 1$

we obtain the following recurrent inequality:

$$\begin{aligned}\left(\mathrm{E}|g_{s,t}|^p\right)^{2/p} &\le \left(c_Q^p \mathrm{E}\Big|\alpha_{t-s} + \sum\nolimits_{s<u\le t} \beta_{t-u,u-s}\zeta_u g_{u,t}\Big|^p\right)^{2/p} \\ &\le c_Q^2 C_p^{2/p}\Big(|\alpha_{t-s}|^p + \mu_p\Big(\sum\nolimits_{s<u\le t}(|\beta_{t-u,u-s}|^p \,\mathrm{E}|g_{u,t}|^p)^{2/p}\Big)^{p/2}\Big)^{2/p} \\ &\le c_Q^2 C_p^{2/p}\Big(|\alpha_{t-s}|^2 + \mu_p^{2/p} \sum_{s<u\le t} \beta^2_{t-u,u-s}(\mathrm{E}|g_{u,t}|^p)^{2/p}\Big).\end{aligned}$$

Iterating the last inequality as in the proof of Theorem 3.3.1 we obtain $(\mathrm{E}|X_t|^p)^{2/p} \le K_{Q,p} < \infty$, with $K_{Q,p}$ given in (3.26). □

Finally, let us discuss the question when X_t of (3.16) satisfies the weak dependence condition in (3.11) for the invariance principle.

Proposition 3.3.4 *Let $\{X_t\}$ satisfy the conditions of Theorem 3.3.1 and $\Omega(2)$ be defined in (3.11). Then*

$$\Omega(2) \;\le\; \sum_{i=0}^{\infty}|\alpha_i| \sum_{k=0}^{\infty} c_Q^{k+1} \sum_{j_1=1}^{\infty} |\beta_{i,j_1}| \cdots \sum_{j_k=1}^{\infty} |\beta_{i+j_1+\cdots+j_{k-1},j_k}|. \tag{3.27}$$

In particular, if the quantity on the r.h.s. of (3.27) is finite, $\{X_t\}$ satisfies the functional central limit theorem in (3.12).

Proof follows from (3.24) and the inequality $|\sum x_i|^{1/2} \le \sum |x_i|^{1/2}$. □

3.4 Examples

Example 3.4.1 (Finitely dependent projective equations) Consider equation (3.16), where $\alpha_i = \beta_{i,j} = 0$ for all $i > m$ and some $m \ge 0$. Since $Q(0) = 0$, the corresponding equation writes as

$$X_t \;=\; \mu + \sum_{t-m<s\le t} \zeta_s Q\Big(\alpha_{t-s} + \sum_{s<u\le t} \beta_{t-u,u-s}\left(\mathrm{E}_{[u,t]}X_t - \mathrm{E}_{[u+1,t]}X_t\right)\Big), \tag{3.28}$$

where the r.h.s. is $\mathcal{F}_{[t-m+1,t]}$-measurable. In particular, $\{X_t\}$ of (3.28) is an m-dependent process. We may ask if the above process can be represented as a moving-average of length m w.r.t. to some i.i.d. innovations? In other words, if there exists an i.i.d. standardized sequence $\{\eta_s, s \in \mathbb{Z}\}$ and coefficients $c_j, 0 \le j < m$ such that

$$X_t = \sum_{t-m<s\le t} c_{t-s}\eta_s, \qquad t \in \mathbb{Z}. \tag{3.29}$$

To construct a negative counter-example to the above question, consider the simple case of (3.28) with $m = 2$, $\mu = 0$, $\alpha_1 = 0, \beta_{0,1} = 1, Q(\alpha_0) = 1$:

$$X_t \;=\; \zeta_t Q(\alpha_0) + \zeta_{t-1} Q(\alpha_1 + \beta_{0,1}\mathrm{E}_{[t,t]}X_t) = \zeta_t + \zeta_{t-1}Q(\zeta_t). \tag{3.30}$$

Assume that $\mathrm{E}Q(\zeta_t) = 0$. Then $\mathrm{E}X_tX_{t-1} = 0, \mathrm{E}X_t^2 = 1 + \mathrm{E}Q^2(\zeta_0)$. On the other hand, from (3.29) with $m = 2$ we obtain $0 = \mathrm{E}X_tX_{t-1} = c_0c_1$, implying that $\{X_t\}$ is an i.i.d. sequence.

Let us show that the last conclusion contradicts the form of X_t in (3.30) under general assumptions on Q and the distribution of $\zeta = \zeta_0$. Assume that ζ is symmetric, $\infty > \mathrm{E}\zeta^4 > (\mathrm{E}\zeta^2)^2 = 1$ and Q is antisymmetric. Then

$$\mathrm{Cov}(X_t^2, X_{t-1}^2) \;=\; \mathrm{E}Q^2(\zeta)\Big\{(\mathrm{E}\zeta^4 - 1) + (\mathrm{E}\zeta^2 Q^2(\zeta) - \mathrm{E}Q^2(\zeta))\Big\}.$$

Assume, in addition, that Q is monotone nondecreasing on $[0,\infty)$. Then $\mathrm{E}\zeta^2Q^2(\zeta) \ge \mathrm{E}\zeta^2\mathrm{E}Q^2(\zeta) = \mathrm{E}Q^2(\zeta)$, implying $\mathrm{Cov}(X_t^2, X_{t-1}^2) > 0$. As a consequence, (3.30) is not a moving average of length 2 in some standardized i.i.d. sequence.

Example 3.4.2 (Linear kernel Q) For linear kernel $Q(x) = c_Q x$, the solution of (3.16) of Theorem 3.3.1 can be written explicitly as $X_t = \mu + \sum_{k=1}^\infty X_t^{(k)}$, where $X_t^{(1)} = c_Q \sum_{i=0}^\infty \alpha_i \zeta_{t-i}$ is a linear process and

$$X_t^{(k+1)} \;=\; c_Q^{k+1} \sum_{i=0}^\infty \alpha_i \sum_{j_1,\dots,j_k=1}^\infty \beta_{i,j_1}\dots\beta_{i+j_1+\dots+j_{k-1},j_k}\zeta_{t-i}\zeta_{t-i-j_1}\dots\zeta_{t-j_1-\dots-j_k}$$

for $k \ge 1$ is a Volterra series of order $k+1$ (see Dedecker et al. (2007), p.22), which are orthogonal in sense that $\mathrm{E}X_t^{(k)}X_s^{(\ell)} = 0$, $t, s \in \mathbb{Z}$, $k, \ell \ge 1$, $k \ne \ell$.

Let $H^2_{(-\infty,t]} \subset L^2_{(-\infty,t]}$ be the subspace spanned by products $1, \zeta_{s_1}, \dots, \zeta_{s_k}$, $s_1 < \dots < s_k \le t, k \ge 1$. Clearly, the above Volterra series $X_t, X_t^{(k)} \in H^2_{(-\infty,t]}$, $\forall t \in \mathbb{Z}$ (corresponding to linear Q) constitute a very special class of projective processes. For example, the process in (3.30) cannot be expanded in such series unless Q is linear. To show the last fact, decompose (3.30) as $X_t = Y_t + Z_t$, where $Y_t := \zeta_t + \alpha\zeta_{t-1}\zeta_t \in H^2_{(-\infty,t]}$, $\alpha := \mathrm{E}\zeta Q(\alpha)$ and $Z_t := \zeta_{t-1}(Q(\zeta_t) - \alpha\zeta_t)$ is orthogonal to $H^2_{(-\infty,t]}$, $Z_t \ne 0$, hence $X_t \notin H^2_{(-\infty,t]}$.

Example 3.4.3 (The LARCH model) The Linear ARCH (LARCH) model, introduced by Robinson (1991) (see also Giraitis et al. (2000), Giraitis et al. (2004), Giraitis et al. (2009)), Giraitis and Surgailis (2002)), is defined by the equations

$$r_t = \sigma_t \zeta_t, \quad \sigma_t = \alpha + \sum_{j=1}^{\infty} \beta_j r_{t-j}, \tag{3.31}$$

where $\{\zeta_t\}$ is a standardized i.i.d. sequence, and the coefficients β_j satisfy $B := \left\{ \sum_{j=1}^{\infty} \beta_j^2 \right\}^{1/2} < \infty$. It is well-known (Giraitis and Surgailis (2002)) that a second order strictly stationary solution $\{r_t\}$ to (3.31) exists if and only if

$$B \ < \ 1, \tag{3.32}$$

in which case it can be represented by the convergent orthogonal Volterra series

$$r_t = \sigma_t \zeta_t, \quad \sigma_t = \alpha \Big(1 + \sum_{k=1}^{\infty} \sum_{j_1, \ldots, j_k = 1}^{\infty} \beta_{j_1} \ldots \beta_{j_k} \zeta_{t-j_1} \ldots \zeta_{t-j_1-\cdots-j_k} \Big).$$

Clearly, the last series is a particular case of the Volterra series of the previous example. We conclude that under the condition (3.32), the volatility process $\{X_t = \sigma_t\}$ of the LARCH model satisfies the projective equation (3.18) with linear function $Q(x) = x$ and $\alpha_j = \alpha \beta_j$. Note that (3.32) coincides with the condition $c_Q^2 B^2 < 1$ of Proposition 3.3.2 (ii) for the existence of solution of (3.18).

From Proposition 3.3.3 the following new result about the existence of higher order moments of the LARCH model is derived.

Corollary 3.4.1 *Assume that*

$$C_p^{1/p} \mu_p^{1/p} B < 1, \tag{3.33}$$

where $\mu_p = \mathrm{E}|\zeta_0|^p$ *and* C_p *is the absolute constant from Proposition 3.2.1,* $p \geq 2$. *Then* $\mathrm{E}|r_t|^p = \mu_p \mathrm{E}|\sigma_t|^p < \infty$. *Moreover,*

$$\mathrm{E}|\sigma_t|^p \ \leq \ \frac{\alpha^2 C_p^{4/p} \mu_p^{2/p} B^2}{1 - C_p^{2/p} \mu_p^{2/p} B^2}. \tag{3.34}$$

Proof follows from Proposition 3.3.3 and the easy fact that for the LARCH model, $K_{Q,p}$ of (3.26) coincides with the r.h.s. of (3.34). □

Condition (3.33) can be compared with the sufficient condition for $\mathrm{E}|r_t|^p < \infty, p = 2, 4, \ldots$ in Giraitis et al. (2000), Lemma 3.1:

$$(2^p - p - 1)^{1/2} \mu_p^{1/p} B < 1. \tag{3.35}$$

Although the best constant C_p in the Rosenthal's inequality is not known, (3.33) seems much weaker than (3.35), especially when p is large. See, e.g. Hitchenko (1990), where it is shown that $C_p^{1/p} = O(p/\log p)$, $p \to \infty$.

Example 3.4.4 (Projective "threshold" equations) Consider projective equation

$$X_t \;\; = \;\; \zeta_t + \sum_{j=1}^{p} \zeta_{t-j} Q\Big(\mathrm{E}_{[t-j+1,t]} X_t\Big), \tag{3.36}$$

where $1 \le p < \infty$ and Q is a bounded measurable function with $Q(0) = 1$. If Q is a step function: $Q(x) = \sum_{k=1}^{q} c_k \mathbf{1}(x \in I_k)$, where $\cup_{k=1}^{q} I_k = \mathbb{R}$ is a partition of $\mathbb{R}$ into disjoint intervals $I_k, 1 \le k \le q$, the process in (3.36) follows different "moving-average regimes" in the regions $\mathrm{E}_{[t-j+1,t]} X_t \in I_k, 1 \le j \le p$ exhibiting a "projective threshold effect". See Figure 3.1, where the top graph shows a trajectory having a single threshold at $x = 0$ and the bottom graph a trajectory with two threshold points at $x = 0$ and $x = 2$.

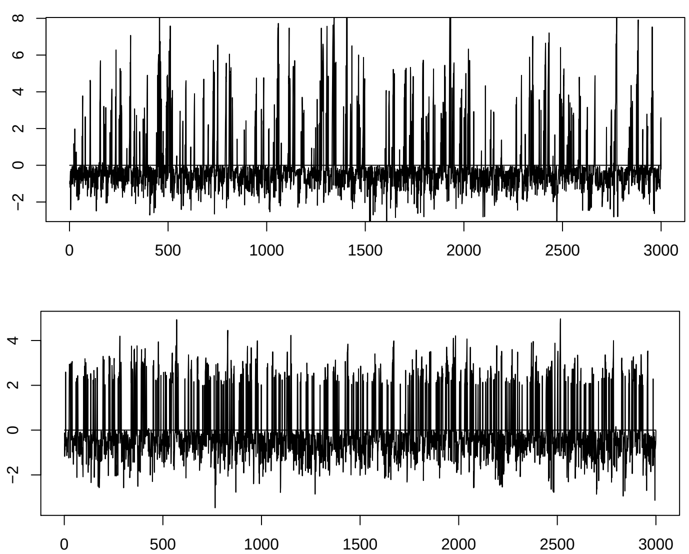

Figure 3.1: Trajectories of solutions of (3.36), $p = 10$. Top: $Q(x) = \mathbf{1}(x > 0)$, bottom: $Q(x) = \mathbf{1}(0 < x < 2)$.

3.5 Simulations

Solutions of projective equations can be easily simulated using a truncated expansion $X_t^{(M)} = \sum_{t-M\le s\le t} g_{s,t}\zeta_s$ instead of infinite series in (3.1). We chose the truncation level M equal to the sample size $M = n = 3000$ in the subsequent simulations. The coefficients $g_{s,t}$ of projective equations are computed very fast from recurrent formula (3.22) and simulated values $\zeta_s, -M \le s \le n$. The innovations were taken standard normal. For better comparisons, we used the same sequence $\zeta_s, -M \le s \le n$ in all simulations.

Stationary solution of equation (3.18) was simulated for three different choices of Q and two choices of coefficients α_j, β_j. The first choice of coefficients is $\alpha_j = 0.5^j, \beta_j = c\,0.9^j$ and corresponds to a short memory process $\{X_t\}$. The second choice is $\alpha_j = \Gamma(d+j)/\Gamma(d)\Gamma(j+1)$, $\beta_j = c\alpha_j$ with $d = 0.4$ corresponds to a long memory process $\{X_t\}$ with coefficients as in ARFIMA$(0, d, 0)$. The value of $c > 0$ was chosen so that $c_Q^2 B^2 = 0.9 < 1$. The latter condition guarantees the existence of a stationary solution of (3.18), see Proposition 3.3.2.

The simulated trajectories and (smoothed) histograms of their marginal densities strongly depend on the kernel Q. We used the following functions:

$$Q_1(x) = x, \qquad Q_2(x) = \max(0, x), \qquad Q_3(x) = \begin{cases} x, & x \in [0, 1], \\ 2 - x, & x \in [1, 2], \\ 0, & \text{otherwise.} \end{cases} \tag{3.37}$$

Clearly, $Q_i, i = 1, 2, 3$ in (3.37) satisfy (3.19) with $c_Q = 1$ and the Lipschitz condition (3.45). Note that Q_3 is bounded and supported in the compact interval $[0, 2]$ while Q_1, Q_2 are unbounded, the latter being bounded from below. Also note that for $\beta_j \equiv 0$ and the choice of α_j as above, the projective process $\{X_t\}$ of (3.18) agrees with AR(.5) for $\alpha_j = 0.5^j$ and with ARFIMA$(0, 0.4, 0)$ for $\alpha_j = \Gamma(d+j)/\Gamma(d)\Gamma(j+1)$ in all three cases in (3.37)

A general impression from our simulations is that in all cases of Q in (3.37), the coefficients α_j account for the persistence and β_j for the clustering of the process. We observe that as β_j's increase, the process becomes more asymmetric and its empirical density diverges from the normal density (plotted in red in Figures 3.2-3.4 with parameters equal to the empirical mean and variance of the simulated series). In the case of unbounded $Q = Q_1, Q_2$ and long memory ARFIMA coefficients, the marginal distribution seems strongly skewed to the left and having a very light left tail and a much heavier right tail. On the other hand, in the case of geometric coefficients,

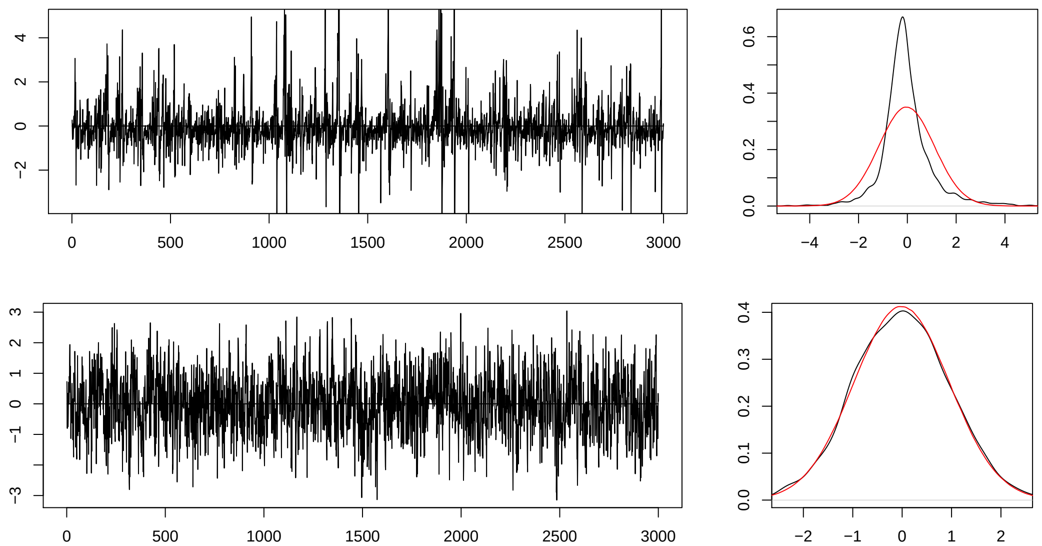

Figure 3.2: Trajectories and (smoothed) histograms of solutions of projective equation (3.18) with $Q(x) = Q_1(x) = x$. Top: $\alpha_j = (.5)^j, \beta_j = c(.9)^j$, bottom: $\alpha_j = (.5)^j, \beta_j = 0$.

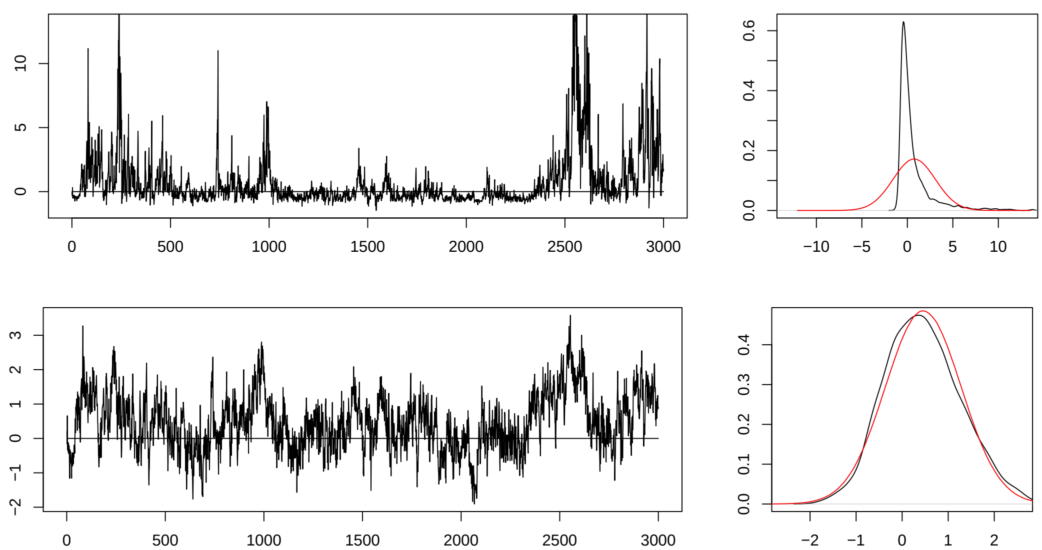

Figure 3.3: Trajectories and (smoothed) histograms of solutions of projective equation (3.18) with $Q(x) = Q_2(x) = \max(x, 0)$. Top: $\alpha_j = \mathrm{ARFIMA}(0, 0.4, 0), \beta_j = c\,\alpha_j$, bottom: $\alpha_j = \mathrm{ARFIMA}(0, 0.4, 0), \beta_j = 0$.

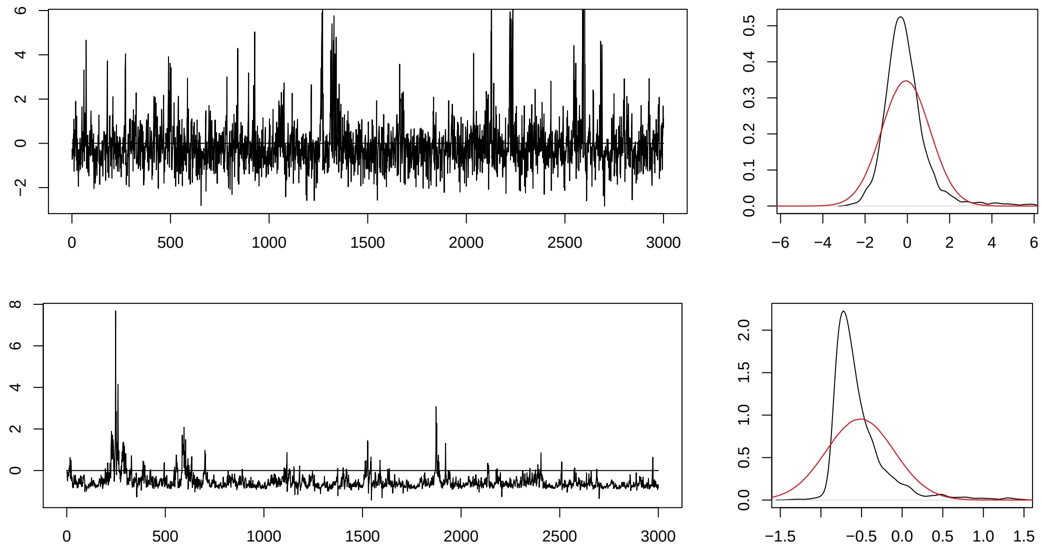

Figure 3.4: Trajectories and (smoothed) histograms of solutions of projective equation (3.18) with $Q(x) = Q_3(x) =$ the "triangle function" in (3.37). Top: $\alpha_j = (.5)^j, \beta_j = c(.9)^j$, bottom: $\alpha_j = \mathrm{ARFIMA}(0, 0.4, 0), \beta_j = c\,\alpha_j$

the density for $Q = Q_1, Q_2$ seems rather symmetric although heavy tailed. Case of $Q = Q_3$ corresponding to bounded Q seems to result in asymmetric distribution with light tails.

3.6 Modifications

Equation (3.16) can be modified in several ways. The first modification is obtained by taking the α_{t-s}'s "outside of Q":

$$X_t = \mu + \sum_{s \le t} \zeta_s \alpha_{t-s} Q\Big(\sum_{s<u\le t} \beta_{t-u,u-s} \left(\mathrm{E}_{[u,t]} X_t - \mathrm{E}_{[u+1,t]} X_t \right) \Big), \qquad (3.38)$$

where $\alpha_i, \beta_{i,j}, Q$ satisfy similar conditions as in (3.16). However, note that (3.19) implies $Q(0) = 0$ in which case (3.38) has a trivial solution $X_t \equiv \mu$. To avoid the last eventuality, condition (3.19) must be changed. Instead, we shall assume that Q is a measurable function satisfying

$$Q(x)^2 \le c_0^2 + c_1^2 x^2, \quad x \in \mathbb{R} \qquad (3.39)$$

for some $c_0, c_1 \ge 0$. Denote

$$\tilde K_Q \ := \ c_0^2 \sum_{k=0}^\infty c_1^{2k} \sum_{i=0}^\infty \alpha_i^2 \sum_{j_1=1}^\infty \alpha_{i+j_1}^2 \beta_{i,j_1}^2 \cdots \sum_{j_k=1}^\infty \alpha^2_{i+j_1+\cdots+j_k} \beta^2_{i+j_1+\cdots+j_{k-1},j_k}.$$

Proposition 3.6.1 can be proved similarly to Theorem 3.3.1 and its proof is omitted.

Proposition 3.6.1 *(i) Assume condition (3.39) and*

$$\tilde K_Q \ < \ \infty. \tag{3.40}$$

Then there exists a unique solution $\{X_t\}$ of (3.38), which is written as a projective moving average of (3.7) with coefficients $g_{t-k,t}$ recursively defined as

$$g_{t-k,t} \ := \ \alpha_k Q\Big(\sum_{i=0}^{k-1} \beta_{i,k-i} \zeta_{t-i} g_{t-i,t}\Big), \qquad k = 1, 2, \dots, \quad g_{t,t} := \alpha_0 Q(0).$$

(ii) In the case of linear function $Q(x) = c_0 + c_1 x$, condition (3.40) is also necessary for the existence of a solution of (3.38).

Remark 3.6.1 Let $A_k^2 := \sum_{i=k}^\infty \alpha_i^2$ and $|\beta_{i,j}| \le \bar\beta$. Then

$$\tilde K_Q \ \le \ c_0^2 \sum_{k=0}^\infty (c_1\bar\beta)^{2k} \sum_{i=0}^\infty \alpha_i^2 \sum_{j_1=1}^\infty \alpha^2_{i+j_1} \cdots \sum_{j_k=1}^\infty \alpha^2_{i+j_1+\cdots+j_k} \ \le \ c_0^2 \sum_{k=0}^\infty (c_1\bar\beta)^{2k} A_0^2 A_1^2 \dots A_k^2.$$

Hence, $A^2 = A_0^2 < \infty$ and $\bar\beta < \infty$ imply $\tilde K_Q < \infty$, for any $c_0, c_1, \bar\beta$; see the proof of Proposition 3.3.2.

Projective stochastic equations (3.16) and (3.38) can be further modified by including projections of lagged variables. Consider the following extension of (3.16):

$$X_t = \mu + \sum_{s \le t} \zeta_s Q\Big(\alpha_{t-s} + \sum_{u=s+1}^{t-1} \beta_{t-1-u,u-s} \left(\mathrm{E}_{[u,t-1]} X_{t-1} - \mathrm{E}_{[u+1,t-1]}\right) X_{t-1}\Big), \tag{3.41}$$

where $\alpha_i, \beta_{i,j}$, Q are the same as in (3.16) and the only new feature is that t is replaced by $t-1$ in the inner sum on the r.h.s. of the equation. This fact allows to study *nonstationary* solutions of (3.41) with a given projective initial condition $X_t = X_t^0, t \le 0$ and the convergence of X_t to the equilibrium as $t \to \infty$; however, we shall not pursue this topic in the present paper. The following proposition is a simple extension of Theorem 3.3.1 and its proof is omitted.

Proposition 3.6.2 *Let $\alpha_i, \beta_{i,j}, Q$ satisfy the conditions of Theorem 3.3.1, including (3.19) and (3.21). Then there exists a unique solution $\{X_t\}$ of (3.41), which is written as a projective moving average of (3.7) with coefficients $g_{t-k,t}$ recursively defined as $g_{t-k,t} := Q(\alpha_k), k = 0, 1$ and*

$$g_{t-k,t} \quad := \quad Q\Big(\alpha_k + \sum_{i=0}^{k-2} \beta_{i,k-1-i}\zeta_{t-1-i}g_{t-1-i,t-1}\Big), \qquad k \geq 2.$$

Finally, consider a projective equation (3.13) with $\mu_t \equiv 0$ and kernels $Q_{s,t} = Q_{t-s}(x_{s+1,t-1}, \ldots, x_{s+1,s})$ depending on $t-s$ real variables, where $Q_0 = 1$ and

$$Q_j(x_1, \ldots, x_j) \quad = \quad \frac{d(x_1)}{1} \cdot \frac{d(x_2)+1}{2} \cdot \frac{d(x_3)+2}{3} \cdots \frac{d(x_j)+j-1}{j}, \tag{3.42}$$

$j \geq 1$, where $d(x), x \in \mathbb{R}$ is a measurable function taking values in the interval $(-1/2, 1/2)$. More explicitly,

$$X_t \quad = \quad \sum_{j=0}^{\infty} Q_j\Big(\mathrm{E}_{[t-j+1,t-1]}X_{t-1}, \mathrm{E}_{[t-j+1,t-2]}X_{t-2}, \ldots, \mathrm{E}_{[t-j+1,t-j]}X_{t-j}\Big)\zeta_{t-j}, \tag{3.43}$$

where $\mathrm{E}_{[t-j+1,t-j]}X_{t-j} = \mathrm{E}X_t = 0$. Note that when $d(x) = d$ is constant, $\{X_t\}$ (3.43) is a stationary ARFIMA$(0, d, 0)$ process. Time-varying fractionally integrated processes with deterministic coefficients of the form (3.42) were studied in Philippe et al. (2006), Philippe et al. (2008). We expect that (3.43) feature a "random" memory intensity depending on the values of the process. A rigorous study of long memory properties of this model does not seem easy. On the other hand, solvability of (3.43) can be established similarly to the previous cases (see below).

Proposition 3.6.3 *Let $d(x)$ be a measurable function taking values in $(-1/2, 1/2)$ and such that $\sup_{x\in\mathbb{R}} d(x) \leq \bar{d}$, where $\bar{d} \in (0, 1/2)$. Then there exists a unique stationary solution $\{X_t\}$ of (3.43), which is written as a projective moving average of (3.7) with coefficients $g_{s,t}$ recursively defined as $g_{t,t} := 1$ and*

$$g_{s,t} \quad := \quad Q_{t-s}\Big(\sum_{s<u\leq t-1} \zeta_u g_{u,t-1}, \sum_{s<u\leq t-2} \zeta_u g_{u,t-2}, \ldots, 0\Big), \qquad s < t, \tag{3.44}$$

with Q_{t-s} defined at (3.42).

Proof. Note that $\sup_{x_1,\ldots,x_j\in\mathbb{R}} |Q_j(x_1, \ldots, x_j)| \leq \Gamma(\bar{d}+j)/\Gamma(\bar{d})\Gamma(j) =: \psi_j$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Therefore the $g_{s,t}$'s in (3.44) satisfy $\sum_{s\leq t} \mathrm{E}g_{s,t}^2 < \infty$ for any $t \in \mathbb{Z}$. The rest of the proof is analogous as the case of Theorem 3.3.1. $\square$

3.7 Long memory

In this section we study long memory properties (the decay of covariance and partial sums' limits) of projective equations (3.16) and (3.38) in the case when the coefficients α_j's decay slowly as $j^{d-1}, 0 < d < 1/2$.

Theorem 3.7.1 *Let $\{X_t\}$ be the solution of projective equation (3.16) satisfying the conditions of Theorem 3.3.1 and $\mu = \mathrm{E}X_t = 0$. Assume, in addition, that Q is a Lipschitz function, viz., there exists a constant $c_L > 0$ such that*

$$|Q(x) - Q(y)| \; < \; c_L|x-y|, \qquad x, y \in \mathbb{R} \tag{3.45}$$

and that there exist $\kappa > 0$ and $0 < d < 1/2$ such that

$$b_j \; := \; Q(\alpha_j) \; \sim \; \kappa\, j^{d-1}, \qquad j \to \infty \tag{3.46}$$

and

$$\bar{\beta}_j \; := \; \max_{0 \le i < j} |\beta_{i,j-i}| \; = \; o(b_j), \qquad j \to \infty. \tag{3.47}$$

Then, as $t \to \infty$

$$\mathrm{E}X_0X_t \; \sim \; \sum_{k=0}^{\infty} b_k b_{t+k} \; \sim \; \kappa_d^2 t^{2d-1}, \tag{3.48}$$

where $\kappa_d^2 := \kappa^2 B(d, 1-d)$ and $B(d, 1-d)$ is beta function. Moreover, as $n \to \infty$

$$n^{-1/2-d} \sum_{t=1}^{[n\tau]} X_t \; \longrightarrow_{D[0,1]} \; c_{\kappa,d} B_H(\tau), \tag{3.49}$$

where B_H is a fractional Brownian motion with parameter $H = d + (1/2)$ and variance $\mathrm{E}B_H^2(t) = t^{2H}$ and $c_{\kappa,d}^2 := \frac{\kappa^2 B(d,1-d)}{d(1+2d)}$.

Proof. Let us note that the statements (3.48) and (3.49) are well-known when $\beta_{i,j} \equiv 0$, in which case X_t coincides with the linear process $Y_t := \sum_{s \le t} b_{t-s}\zeta_s$, see, e.g., Giraitis et al. (2012), Proposition 3.2.1 and Corollary 4.4.1.

The natural idea of the proof is to approximate $\{X_t\}$ by the linear process $\{Y_t\}$.

For $t \geq 0, k \geq 0$, denote

$$\begin{aligned} r^X_t &:= \mathrm{E}X_0X_t \;=\; \sum_{s\le 0}\mathrm{E}[g_{s,0}\, g_{s,t}], \qquad r^Y_t \;:=\; \mathrm{E}Y_0Y_t \;=\; \sum_{s\le 0} b_{-s}b_{t-s}, \\ \varphi_{t-k,t} &:= g_{t-k,t} - b_k \;=\; Q\Big(\alpha_k + \sum_{i=0}^{k-1}\beta_{i,k-i}\zeta_{t-i}g_{t-i,t}\Big) - Q(\alpha_k). \end{aligned}$$

Then

$$\begin{aligned} r^X_t - r^Y_t &= \sum_{s\le 0}\mathrm{E}[(b_{-s}+\varphi_{s,0})(b_{t-s}+\varphi_{s,t}) - b_{-s}b_{t-s}] \\ &= \sum_{s\le 0} b_{-s}\mathrm{E}[\varphi_{s,t}] + \sum_{s\le 0} b_{t-s}\mathrm{E}[\varphi_{s,0}] + \sum_{s\le 0}\mathrm{E}[\varphi_{s,0}\,\varphi_{s,t}] \;=:\; \sum_{i=1}^{3}\rho_{i,t}. \end{aligned}$$

Using (3.45) we obtain

$$\begin{aligned} |\mathrm{E}\varphi_{t-k,t}|^2 \;\le\; \mathrm{E}\varphi^2_{t-k,t} &\le c_L^2\mathrm{E}\Big(\sum_{i=0}^{k-1}\beta_{i,k-i}\zeta_{t-i}g_{t-i,t}\Big)^2 \\ &= c_L^2\Big(\sum_{i=0}^{k-1}\beta^2_{i,k-i}\mathrm{E}g^2_{t-i,t}\Big) \\ &\le \bar\beta_k^2 c_L^2\Big(\sum_{i=0}^{\infty}\mathrm{E}g^2_{t-i,t}\Big) \\ &\le \bar\beta_k^2 c_L^2 K_Q. \end{aligned}$$

This and condition (3.47) imply that

$$|\mathrm{E}\varphi_{t-k,t}| + \mathrm{E}^{1/2}\varphi^2_{t-k,t} \;\le\; \delta_k k^{d-1}, \qquad \forall\, t,k\ge 0,$$

where $\delta_k \to 0\,(k\to\infty)$. Therefore for any $t \geq 1$

$$\begin{aligned} |\rho_{1,t}| &\le C\sum_{k=1}^{\infty}k^{d-1}(t+k)^{d-1}\delta_{t+k} \;\le\; C\delta'_t t^{2d-1}, \\ |\rho_{2,t}| &\le C\sum_{k=1}^{\infty}k^{d-1}\delta_k(t+k)^{d-1} \;\le\; C\delta'_t t^{2d-1}, \\ |\rho_{3,t}| &\le \sum_{s\le t}\mathrm{E}^{1/2}[\varphi^2_{s,0}]\,\mathrm{E}^{1/2}[\varphi^2_{s,t}] \;\le\; C\sum_{k=1}^{\infty}k^{d-1}(t+k)^{d-1}\delta_k\delta_{t+k} \;\le\; C\delta'_t t^{2d-1}, \end{aligned}$$

where $\delta'_k \to 0$ $(k\to\infty)$. This proves (3.48).

To show (3.49), consider $Z_t := X_t - Y_t = \sum_{u\le t}\varphi_{u,t}\zeta_u, t\in\mathbb{Z}$. By stationarity of

$\{Z_t\}$, for any $s \le t$ we have

$$\mathrm{Cov}(Z_t, Z_s) = \sum_{u\le 0} \mathrm{E}[\varphi_{u,0}\, \varphi_{u,t-s}] \le \sum_{u\le 0} \mathrm{E}^{1/2}[\varphi^2_{u,0}]\, \mathrm{E}^{1/2}[\varphi^2_{u,t-s}] = o((t-s)^{2d-1}),$$

see above, and therefore $\mathrm{E}\Big(\sum_{t=1}^n Z_t\Big)^2 = o(n^{2d+1})$, implying

$$n^{-d-(1/2)} \sum_{t=1}^{[n\tau]} X_t \;=\; n^{-d-(1/2)} \sum_{t=1}^{[n\tau]} Y_t + o_p(1).$$

Therefore partial sums of $\{X_t\}$ and $\{Y_t\}$ tend to the same limit $c_{\kappa,d} B_H(\tau)$, in the sense of weak convergence of finite dimensional distributions. The tightness in $D[0,1]$ follows from (3.48) and the Kolmogorov criterion. Theorem 3.7.1 is proved. $\square$

A similar but somewhat different approximation by a linear process applies in the case of projective equations of (3.38). Let us discuss a special case of $\beta_{i,j}$:

$$\beta_{i,j} = 1, \qquad \text{for all } i = 0, 1, \ldots,\ j = 1, 2, \ldots. \tag{3.50}$$

Note that for such $\beta_{i,j}$, $\sum_{s<u\le t} \beta_{t-u,u-s}(\mathrm{E}_{[u,t]} - \mathrm{E}_{[u+1,t]})X_t = \mathrm{E}_{[s+1,t]} X_t$, $s < t$ and the corresponding projective equation (3.38) with $\mu = 0, \alpha_i = b_i$ coincides with (3.5). Recall that for bounded $\beta_{i,j}$'s as in (3.50), condition (3.39) on Q together with $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$ guarantee the existence of the stationary solution $\{X_t\}$ (see Remark 3.6.1). We shall also need the following additional condition:

$$\mathrm{E}\Big(Q(\mathrm{E}_{[s,0]} X_0) - Q(X_0)\Big)^2 \;\to\; 0, \qquad \text{as} \quad s \to -\infty. \tag{3.51}$$

Since $\mathrm{E}\Big(\mathrm{E}_{[s,0]} X_0 - X_0\Big)^2 \to 0$, $s \to -\infty$, so (3.51) is satisfied if Q is Lipschitz, but otherwise conditions (3.51) and (3.39) allow Q to be even discontinuous. Denote

$$c^2_{Q,d} := \Big(\mathrm{E}[Q(X_0)]\Big)^2 B(d, 1-d).$$

Theorem 3.7.2 *Let $\{X_t\}$ be the solution of projective equation (3.38) with $\mu = 0$, $\beta_{i,j}$ as in (3.50), Q satisfying (3.39) and*

$$\alpha_k \;\sim\; k^{d-1}, \qquad k \to \infty, \qquad \exists\;\; 0 < d < 1/2. \tag{3.52}$$

In addition, let (3.51) hold. Then

$$\mathrm{E} X_0 X_t \;\sim\; c^2_{Q,d} t^{2d-1}, \qquad t \to \infty \tag{3.53}$$

and

$$n^{-1/2-d}\sum_{t=1}^{[n\tau]} X_t \quad \longrightarrow_{D[0,1]} \quad c'_{Q,d}B_H(\tau), \qquad c'_{Q,d} := c_{Q,d}/(d(1+2d)^{1/2}. \quad (3.54)$$

Proof. Similarly as in the proof of the previous theorem, let $Y_t := \sum_{s\le t} b_{t-s}\zeta_s$, $b_k := \alpha_k \mathrm{E}[Q(X_0)]$, $r^X_t := \mathrm{E}X_0X_t$, $r^Y_t := \mathrm{E}Y_0Y_t$, $t \ge 0$. Relation (3.53) follows from

$$r^X_t - r^Y_t \ = \ o(t^{2d-1}), \qquad t \to \infty. \quad (3.55)$$

We have $X_t = \sum_{s\le t} g_{s,t}\zeta_s$, $g_{s,t} = \alpha_{t-s}Q(\mathrm{E}_{[s+1,t]}X_t)$, $\mathrm{E}X^2_t = \sum_{s\le t}\mathrm{E}g^2_{s,t} < \infty$ and $\mathrm{E}[Q(\mathrm{E}_{[s+1,t]}X_t)^2] \le c_0^2 + c_1^2\mathrm{E}(\mathrm{E}_{[s+1,t]}X_t)^2 \le c_0^2 + c_1^2\mathrm{E}X^2_t < C$. Decompose $r^X_t = r^X_{1,t} + r^X_{2,t}$, where

$$r^X_{1,t} \ := \ \sum_{s\le 0}\alpha_s\alpha_{t+s}\mathrm{E}[Q(\mathrm{E}_{[s+1,0]}X_0)]\,\mathrm{E}[Q(\mathrm{E}_{[1,t]}X_t)], \qquad r^X_{2,t} \ := \ \sum_{s\le 0}\alpha_s\alpha_{t+s}\gamma_{s,t},$$

and where

$$|\gamma_{s,t}| \ := \ \Big|\mathrm{E}\Big[Q(\mathrm{E}_{[s+1,0]}X_0)\Big\{Q(\mathrm{E}_{[s+1,t]}X_t) - Q(\mathrm{E}_{[1,t]}X_t)\Big\}\Big]\Big| \ \le \ \tilde\gamma^{1/2}_{1,s}\tilde\gamma^{1/2}_{2,s,t}.$$

Here, $\tilde\gamma_{1,s} := \mathrm{E}[Q^2(\mathrm{E}_{[s+1,0]}X_0)] \le C$, see above, while

$$\begin{aligned} |\tilde\gamma_{2,s,t}| \ &:= \ \mathrm{E}\Big[\Big(Q(\mathrm{E}_{[s+1,t]}X_t) - Q(\mathrm{E}_{[1,t]}X_t)\Big)^2\Big] \\ &= \ \mathrm{E}\Big[\Big(Q(\mathrm{E}_{[s+1-t,0]}X_0) - Q(\mathrm{E}_{[1-t,0]}X_0)\Big)^2\Big] \ \to \ 0, \qquad t\to\infty \quad (3.56)\end{aligned}$$

uniformly in $s \le 0$, according to (3.51). Hence and from (3.52) it follows that

$$|r^X_{2,t}| \ = \ o(t^{2d-1}), \qquad t \to \infty. \quad (3.57)$$

Accordingly, it suffices to prove (3.55) with r^X_t replaced by $r^X_{1,t}$. We have

$$r^X_{1,t} \ = \ r^Y_t + \sum_{s\le 0}\alpha_s\alpha_{t+s}\varphi_{1,s,t} + \sum_{s\le 0}\alpha_s\alpha_{t+s}\varphi_{2,s,t} + \sum_{s\le 0}\alpha_s\alpha_{t+s}\varphi_{3,s,t},$$

where the "remainders" $\varphi_{1,s,t} := \mathrm{E}[Q(X_0)]\Big\{\mathrm{E}[Q(\mathrm{E}_{[s+1,0]}X_0)] - \mathrm{E}[Q(X_0)]\Big\}$, $\varphi_{2,s,t} := \mathrm{E}[Q(X_0)]\Big\{\mathrm{E}[Q(\mathrm{E}_{[1-t,0]}X_0)] - \mathrm{E}[Q(X_0)]\Big\}$ and $\varphi_{3,s,t} := \Big(\mathrm{E}[Q(\mathrm{E}_{[s+1,0]}X_0)] - \mathrm{E}[Q(X_0)]\Big) \times\Big(\mathrm{E}[Q(\mathrm{E}_{[1-t,0]}X_0)] - \mathrm{E}[Q(X_0)]\Big)$ can be estimated similarly to (3.56), leading to the asymptotics of (3.57) for each of the three sums in the above decomposition of $r^X_{1,t}$. This proves (3.53).

Let us prove (3.54). Consider the convergence of one-dimensional distributions for $\tau = 1$, viz.,

$$n^{-d-1/2} S_n^X \ \to\ \mathcal{N}(0, \sigma^2), \qquad \sigma = c'_{Q,d}, \tag{3.58}$$

where $S_n^X := \sum_{t=1}^n X_t$. Then (3.58) follows from

$$\mathrm{E}(S_n^X - S_n^Y)^2 \ = \ o(n^{1+2d}), \tag{3.59}$$

where $S_n^Y := \sum_{t=1}^n Y_t$ and Y_t is as above. We have

$$\begin{aligned} \mathrm{E}(S_n^X - S_n^Y)^2 \ &= \ \mathrm{E}\Big(\sum_{s\le n} \zeta_s \sum_{t=1\vee s}^n \alpha_{t-s} \tilde{Q}_{s,t}\Big)^2 \\ &= \ \sum_{s\le n} \sum_{t_1,t_2=1\vee s}^n \alpha_{t_1-s}\alpha_{t_2-s} \mathrm{E}[\tilde{Q}_{s,t_1}\tilde{Q}_{s,t_2}], \end{aligned} \tag{3.60}$$

where $\tilde{Q}_{s,t} := Q\Big(\mathrm{E}_{[s+1,t]} X_t\Big) - \mathrm{E}[Q(X_0)]$. Let us prove that uniformly in $s \le t_1$

$$\mathrm{E}[\tilde{Q}_{s,t_1}\tilde{Q}_{s,t_2}] \ = \ o(1), \qquad \text{as} \qquad t_2 - t_1 \to \infty. \tag{3.61}$$

We have for $s \le t_1 \le t_2$ that

$$\begin{aligned} \mathrm{E}[\tilde{Q}_{s,t_1}\tilde{Q}_{s,t_2}] \ &= \ \mathrm{E}\Big[\tilde{Q}_{s,t_1}\Big\{Q\Big(\mathrm{E}_{[s+1,t_2]}X_{t_2}\Big) - \mathrm{E}[Q(X_0)]\Big\}\Big] \\ &= \ \mathrm{E}[\tilde{Q}_{s,t_1}]\Big\{\mathrm{E}\Big[Q\Big(\mathrm{E}_{[t_1+1,t_2]}X_{t_2}\Big)\Big] - \mathrm{E}[Q(X_0)]\Big\} \\ &+ \ \mathrm{E}\Big[\tilde{Q}_{s,t_1}\Big\{Q\Big(\mathrm{E}_{[s+1,t_2]}X_{t_2}\Big) - Q\Big(\mathrm{E}_{[t_1+1,t_2]}X_{t_2}\Big)\Big\}\Big] \ =: \ \psi'_{s,t_1,t_2} + \psi''_{s,t_1,t_2}, \end{aligned}$$

where we used the fact that $\tilde{Q}_{s,t_1}$ and $Q\Big(\mathrm{E}_{[t_1+1,t_2]}X_{t_2}\Big)$ are independent. Here, thanks to (3.51), we see that $|\psi'_{s,t_1,t_2}| \le \mathrm{E}^{1/2}[\tilde{Q}^2_{s,t_1}]\,\mathrm{E}^{1/2}\Big[\Big\{Q\Big(\mathrm{E}_{[t_1-t_2+1,0]}X_0\Big) - Q(X_0)\Big\}^2\Big] \le C\mathrm{E}^{1/2}\Big[\Big\{Q\Big(\mathrm{E}_{[t_1-t_2+1,0]}X_0\Big) - Q(X_0)\Big\}^2\Big] \to 0$ uniformly in $s \le t_1 \le t_2$ as $t_2 - t_1 \to \infty$. The same is true for $|\psi''_{s,t_1,t_2}|$ since it is completely analogous to (3.56). This proves (3.61). Next, with (3.60) in mind, split $\mathrm{E}(S_n^X - S_n^Y)^2 =: T_n = T_{1,n} + T_{2,n}$, where

$$T_{1,n} := \sum_{s\le n} \sum_{t_1,t_2=1\vee s}^n \mathbf{1}(|t_1-t_2| > K)\ldots, \qquad T_{2,n} := \sum_{s\le n} \sum_{t_1,t_2=1\vee s}^n \mathbf{1}(|t_1-t_2| \le K)\ldots,$$

where K is a large number. By (3.61), for any $\epsilon > 0$ we can find $K > 0$ such that $\sup_{s\le t_1<t_2:t_2-t_1>K} |\mathrm{E}[\tilde{Q}_{s,t_1}\tilde{Q}_{s,t_2}]| < \epsilon$ and therefore

$$|T_{1,n}| \ < \ \epsilon \sum_{s\le n} \sum_{t_1,t_2=1\vee s}^n |\alpha_{t_1-s}\alpha_{t_2-s}| \ \le \ C\epsilon \sum_{t_1,t_2=1}^n |\bar{r}_{t_1-t_2}| \ \le \ C\epsilon n^{1+2d}$$

holds for all $n > 1$ large enough, where $\bar{r}_t := \sum_{i=0}^{\infty} |\alpha_i \alpha_{t+i}| = O(t^{2d-1})$ in view of (3.52). On the other hand, $|T_{2,n}| \le CKn = o(n^{1+2d})$ for any $K < \infty$ fixed. Then (3.59) follows, implying the finite-dimensional convergence in (3.54). The tightness in (3.54) follows from (3.53) and the Kolmogorov criterion, similarly as in the proof of Theorem 3.7.1. Theorem 3.7.2 is proved. $\square$

Remark 3.7.1 Shao and Wu (2006) discussed partial sums limits of fractionally integrated nonlinear processes $Y_t = (1-L)^{-d} u_t$, $t \in \mathbb{Z}$, where $LX_t = X_{t-1}$ is the backward shift, $(1-L)^d = \sum_{j=0}^{\infty} \psi_j(d) L^j$, $d \in (-1,1)$ is the fractional differentiation operator, and $\{u_t\}$ is a causal Bernoulli shift:

$$u_t = F(\ldots, \zeta_{t-1}, \zeta_t), \qquad t \in \mathbb{Z} \tag{3.62}$$

in i.i.d. r.v.'s $\{\zeta_t, t \in \mathbb{Z}\}$. The weak dependence condition on $\{u_t\}$ (3.62), analogous to (3.11) and guaranteeing the weak convergence of normalized partial sums of $\{Y_t\}$ towards a fractional Brownian motion, is written in terms of projections $P_0 u_t = (\mathrm{E}_{[0,t]} - \mathrm{E}_{[1,t]}) u_t$:

$$\Omega(q) \ := \ \sum_{t=1}^{\infty} \|P_0 u_t\|_q \ < \ \infty, \tag{3.63}$$

where $\|\xi\|_q := \mathrm{E}^{1/q} |\xi|^q$ and $q = 2$ for $0 < d < 1/2$; see Theorem. 2.1 in Shao and Wu (2006), also Wu and Min (2005), Wu (2005). The above mentioned papers verify (3.63) for several classes of Bernoulli shifts. It is of interest to verify (3.63) for projective moving averages. For X_t of (3.1) and $0 < d < 1/2$, $u_t := (1-L)^d X_t = \sum_{s \le t} \zeta_s G_{s,t}$ is a well-defined projective moving average with coefficients

$$G_{s,t} := \sum_{s \le v \le t} \psi_{t-v}(d) g_{s,v}, \qquad s \le t,$$

see Proposition 3.2.2. For concreteness, let $g_{s,t} = \psi_{t-s}(-d) Q(\mathrm{E}_{[s+1,t]} X_t)$ as in Theorem 3.7.2 with $\alpha_j = \psi_j(-d)$. We have $\Omega(2) = \sum_{t=1}^{\infty} \|G_{0,t}\|_2$, where

$$\|G_{0,t}\|_2^2 = \mathrm{E}\Big[\sum_{v=0}^{t} \psi_{t-v}(d)\psi_v(-d) Q(\mathrm{E}_{[1,v]} X_v)\Big]^2 \ = \mathrm{E}\Big[\sum_{v=0}^{t-1} \psi_{t-v}(d)\psi_v(-d) Q_{v,t}\Big]^2, \tag{3.64}$$

where $Q_{v,t} := Q(\mathrm{E}_{[1,v]} X_v) - Q(\mathrm{E}_{[1,t]} X_t)$ and we used $\sum_{v=0}^{t} \psi_{t-v}(d)\psi_v(-d) = 0$, $t \ge 1$ in the last equality. Note that $\psi_{t-v}(d)\psi_v(-d) < 0$ have the same sign and $Q_{v,t} \approx Q(X_v) - Q(X_t)$ are not negligible in (3.64). Therefore we conjecture that $\|G_{0,t}\|_2^2 = O\Big(\sum_{v=0}^{t-1} |\psi_{t-v}(d)\psi_v(-d)|\Big)^2 = O(t^{-2(1-d)})$ and hence $\Omega(2) = \infty$ for $0 < d < 1/2$. The above argument suggests that projective moving averages posses a different "memory mechanism" from fractionally integrated processes in Shao and Wu (2006).

Chapter 4

A nonlinear model for long memory conditional heteroscedasticity

Abstract. We discuss a class of conditionally heteroscedastic time series models satisfying the equation $r_t = \zeta_t \sigma_t$, where ζ_t are standardized i.i.d. r.v.'s and the conditional standard deviation σ_t is a nonlinear function Q of inhomogeneous linear combination of past values $r_s, s < t$ with coefficients b_j. The existence of stationary solution r_t with finite pth moment, $0 < p < \infty$ is obtained under some conditions on Q, b_j and the pth moment of ζ_0. Weak dependence properties of r_t are studied, including the invariance principle for partial sums of Lipschitz functions of r_t. In the case when Q is the square root of a quadratic polynomial, we prove that r_t can exhibit a leverage effect and long memory, in the sense that the squared process r_t^2 has long memory autocorrelation and its normalized partial sums process converges to a fractional Brownian motion. The results are extended to a generalized model where the conditional variance satisfies an AR(1) equation $\sigma_t^2 = Q^2\Big(a + \sum_{j=1}^{\infty} b_j r_{t-j}\Big) + \gamma \sigma_{t-1}^2$. We also provide another condition for the existence of higher moments of r_t which does not include the Rosenthal constant. A simulated trajectories and histograms of marginal density of σ_t for different values of γ are presented.

4.1 Introduction

A class of conditionally heteroscedastic ARCH-type processes is defined from a standardized i.i.d. sequence $\{\zeta_t, t \in \mathbb{Z}\}$ as solutions of stochastic equation

$$r_t = \zeta_t \sigma_t, \qquad \sigma_t = V(r_s, s < t), \tag{4.1}$$

where $V(x_1, x_2, \dots)$ is some function of $x_1, x_2, \dots$. The present chapter discusses a class of models (4.1) with V of the form

$$V(x_1, x_2, \dots) = Q(a + \sum_{j=1}^{\infty} b_j x_j), \tag{4.2}$$

where $Q(x), x \in \mathbb{R}$ is a (nonlinear) function of a single real variable $x \in \mathbb{R}$ which may be separated from zero by a positive constant: $Q(x) \geq c > 0$, $x \in \mathbb{R}$. Linear $Q(x) = x$ corresponds to the LARCH model (2.11). Probably, the most interesting nonlinear case of Q in (4.2) is

$$Q(x) = \sqrt{c^2 + x^2}, \tag{4.3}$$

where $c \geq 0$ is a parameter. In the latter case, the model is described by equations

$$r_t \ = \ \zeta_t \sigma_t, \qquad \sigma_t \ = \ \sqrt{c^2 + \Big(a + \sum_{s<t} b_{t-s} r_s\Big)^2}. \tag{4.4}$$

Note that $\sigma_t \geq c \geq 0$ in (4.4) is nonnegative and separated from 0 if $c > 0$. Particular cases of volatility forms in (4.4) are:

$$\begin{aligned}
\sigma_t &= \sqrt{c^2 + (a + b r_{t-1})^2} \qquad (\text{ Engle (1990) asymmetric ARCH(1)}), && (4.5)\\
\sigma_t &= \sqrt{c^2 + \Big(a + \frac{b}{p}\sum_{j=1}^{p} r_{t-j}\Big)^2}, && (4.6)\\
\sigma_t &= \Big|a + \sum_{j=1}^{\infty} b_j r_{t-j}\Big| \qquad (Q(x) = |x|), && (4.7)\\
\sigma_t &= \sqrt{c^2 + (a + b((1-L)^{-d} - 1) r_t)^2}. && (4.8)
\end{aligned}$$

In (4.5)-(4.8), a, b, c are real parameters, $p \geq 1$ an integer, $L x_t = x_{t-1}$ is the backward shift, and $(1 - L)^{-d} x_t = \sum_{j=0}^{\infty} \varphi_j x_{t-j}$, $\varphi_j = \Gamma(d + j)/\Gamma(d)\Gamma(j + 1), \varphi_0 = 1$ is the fractional integration operator, $0 < d < 1/2$. The squared volatility (conditional variance) σ_t^2 in (4.5)-(4.8) and (4.4) is a quadratic form in lagged returns $r_{t-1}, r_{t-2}, \dots$

and hence represent particular cases of Sentana (1995) Quadratic ARCH (QARCH) model with $p = \infty$. It should be noted, however, that the first two conditional moments do not determine the unconditional distribution. Particularly, (4.1) with (4.4) generally is a different process from Sentana (1995) QARCH process, the latter being defined as a solution of a linear random-coefficient equation for $\{r_t\}$ in contrast to the nonlinear equation in (4.1). See also Example 4.2.2 below.

The model in (4.1)-(4.2) can be generalized by including the lagged volatilities from the past, in particular

$$r_t \;=\; \zeta_t \sigma_t, \qquad \sigma_t^2 \;=\; Q^2\Big(a + \sum_{j=1}^{\infty} b_j r_{t-j}\Big) + \gamma \sigma_{t-1}^2, \tag{4.9}$$

where $0 \le \gamma < 1$ is a parameter. The inclusion of lagged σ_{t-1}^2 in (4.9) helps to reduce very sharp peaks and clustering of volatility which occur in trajectory of (4.1)-(4.2) with (4.3) near the threshhold $c > 0$ (see Figure 4.1). The generalization from (4.1)-(4.2) to (4.9) is similar to that from ARCH to GARCH models, see Engle (1982), Bollerslev (1986), particularly, (4.9) with $Q(x)$ of (4.3) and $b_j = 0, j \ge 2$ reduces to the Asymmetric GARCH(1,1) of Engle (1990), see Example 4.6.1.

Let us describe the main results of this chapter. Section 4.2 obtains sufficient conditions on Q, b_j and $|\mu|_p := \mathrm{E}|\zeta_0|^p$ for the existence of stationary solution of (4.1)-(4.2) with finite moment $\mathrm{E}|r_t|^p < \infty$, $p > 0$. We use the fact that the above equations can be reduced to the "nonlinear moving-average" equation

$$X_t = \sum_{s<t} b_{t-s} \zeta_s Q(a + X_s)$$

for linear form $X_t := \sum_{s<t} b_{t-s} r_s$ in (4.2), and vice-versa. Section 4.3 aims at providing weak dependence properties of model (4.1)-(4.2), in particular, the invariance principle for Lipschitz functions of $\{r_t\}$ and $\{X_t\}$, under the assumption that b_j are summable and decay as $j^{-\gamma}$ with $\gamma > 1$. Section 4.4 discusses long memory property of the quadratic model in (4.4). For $b_j \sim \beta j^{d-1}$, $j \to \infty$, $0 < d < 1/2$ as in (4.8), we prove that the squared process $\{r_t^2\}$ has long memory autocorrelations and its normalized partial sums process tend to a fractional Brownian motion with Hurst parameter $H = d + 1/2$ (Theorem 4.4.2). Section 4.5 establishes the leverage effect in spirit of Giraitis et al. (2004), viz., the fact that the "leverage function" $h_j := \mathrm{Cov}(\sigma_t^2, r_{t-j}), j \ge 1$ of model (4.4) takes negative values provided the coefficients a and b_j have opposite signs. Finally, Section 4.6 extends the results of previous sections to a more general class of volatility forms in (4.9) that include lagged volatility from the past σ_{t-1}^2. In addition, another condition for the existence of higher moments of r_t which does not include the Rosenthal constant is obtained in Theorem 4.6.2. Simulated trajectories and histograms of marginal density for different values of parameter γ are presented in Section 4.7.

Notation. In what follows, $C, C(\dots)$ denote generic constants, possibly dependent on the variables in brackets, which may be different at different locations. $a_t \sim b_t$ $(t \to \infty)$ is equivalent to $\lim_{t\to\infty} a_t/b_t = 1$.

4.2 Stationary solution

This section discusses the existence of a stationary solution of (4.1)-(4.2), viz.,

$$r_t = \zeta_t Q\Big(a + \sum_{s<t} b_{t-s} r_s\Big), \qquad t \in \mathbb{Z}. \tag{4.10}$$

Denote

$$X_t \ := \ \sum_{s<t} b_{t-s} r_s. \tag{4.11}$$

Then r_t in (4.10) can be written as $r_t = \zeta_t Q(a + X_t)$ where (4.11) formally satisfies the following equation:

$$X_t \ = \ \sum_{s<t} b_{t-s} \zeta_s Q(a + X_s). \tag{4.12}$$

Below we give rigorous definitions of solutions of (4.10) and (4.12) and a statement (Proposition 4.2.2) justifying (4.12) and the equivalence of (4.10) and (4.12).

In this section we consider a general case of (4.10)-(4.12) when the innovations may have infinite variance. More precisely, we assume that $\{\zeta_t, t \in \mathbb{Z}\}$ are i.i.d. r.v.'s with finite moment $|\mu|_p := \mathrm{E}|\zeta_t|^p < \infty$, $p > 0$. We use the following moment inequality.

Proposition 4.2.1 *Let $\{Y_j, j \ge 1\}$ be a sequence of r.v.'s such that $\mathrm{E}|Y_j|^p < \infty$ for some $p > 0$ and the sum on the r.h.s. of (4.13) converges. If $p > 1$ we additionally assume that $\{Y_j\}$ is a martingale difference sequence: $\mathrm{E}[Y_j|Y_1, \dots, Y_{j-1}] = 0$, $j = 2, 3, \dots$. Then there exists a constant K_p depending only on p and such that*

$$\mathrm{E}\Big|\sum_{j=1}^{\infty} Y_j\Big|^p \ \le \ K_p \begin{cases} \sum_{j=1}^{\infty} \mathrm{E}|Y_j|^p, & 0 < p \le 2, \\ \Big(\sum_{j=1}^{\infty} (\mathrm{E}|Y_j|^p)^{2/p}\Big)^{p/2}, & p > 2. \end{cases} \tag{4.13}$$

Remark 4.2.1 In the sequel, we shall refer to K_p in (4.13) as the Rosenthal constant. For $0 < p \le 1$ and $p = 2$, inequality (4.13) holds with $K_p = 1$, and for $1 < p < 2$, it is known as von Bahr and Esséen inequality, see von Bahr and Esséen (1965), which holds with $K_p = 2$. For $p > 2$, inequality (4.13) is a consequence of the Burkholder and Rosenthal inequality (see Burkholder (1973), Rosenthal (1970), also Giraitis et al. (2012), Lemma 2.5.2). Osękowski (2012) proved that $K_p^{1/p} \le 2^{(3/2)+(1/p)}(\frac{p}{4}+1)^{1/p}\Big(1+\frac{p}{\log(p/2)}\Big)$, in particular, $K_4^{1/4} \le 27.083$. See also Hitchenko (1990).

Let us give some formal definitions. 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 $t \in \mathbb{Z}$). Define

$$B_p := \begin{cases} \sum_{j=1}^\infty |b_j|^p, & 0 < p < 2, \\ \Big(\sum_{j=1}^\infty b_j^2\Big)^{p/2}, & p \ge 2. \end{cases} \tag{4.14}$$

Definition 4.2.1 *Let $p > 0$ be arbitrary.*

(i) By L^p-solution of (4.10) *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 $\sum_{s<t} b_{t-s} r_s$ converges in L^p and* (4.10) *holds.*

(ii) By L^p-solution of (4.12) *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 $\sum_{s<t} b_{t-s}\zeta_s Q(a+X_s)$ converges in L^p and* (4.12) *holds.*

Let $Q(x), x \in \mathbb{R}$ be a Lipschitz function, i.e., there exists $\mathrm{Lip}_Q > 0$ such that

$$|Q(x) - Q(y)| \le \mathrm{Lip}_Q |x - y|, \qquad x, y \in \mathbb{R}. \tag{4.15}$$

Note (4.15) implies the bound

$$Q^2(x) \le c_1^2 + c_2^2 x^2, \qquad x \in \mathbb{R}, \tag{4.16}$$

where $c_1 \ge 0$, $c_2 \ge \mathrm{Lip}_Q$ and c_2 can be chosen arbitrarily close to Lip_Q; in particular, (4.16) holds with $c_2^2 = (1+\epsilon^2)\mathrm{Lip}_Q^2, c_1^2 = Q^2(0)(1+\epsilon^{-2})$, where $\epsilon > 0$ is arbitrarily small.

Proposition 4.2.2 *Let Q be a measurable function satisfying* (4.16) *with some $c_i \ge 0$, $i = 1, 2$ and $\{\zeta_t\}$ 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 $B_p < \infty$.*

(i) Let $\{X_t\}$ be a stationary L^p-solution of (4.12)*. Then $\{r_t := \zeta_t Q(a + X_t)\}$ is a stationary L^p-solution of* (4.10) *and*

$$\mathrm{E}|r_t|^p \ \le \ C(1 + \mathrm{E}|X_t|^p). \tag{4.17}$$

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

$$\mathrm{E}[r_t|\mathcal{F}_{t-1}] = 0, \qquad \mathrm{E}[|r_t|^p|\mathcal{F}_{t-1}] = |\mu|_p \Big| Q(a + \sum_{s<t} b_{t-s} r_s)|^p. \tag{4.18}$$

(ii) Let $\{r_t\}$ be a stationary L^p-solution of (4.10)*. Then $\{X_t\}$ in* (4.11) *is a stationary L^p-solution of* (4.12) *such that*

$$\mathrm{E}|X_t|^p \ \le \ C\mathrm{E}|r_t|^p. \tag{4.19}$$

Moreover, for $p \ge 2$

$$\mathrm{E}[X_t X_0] \ = \ \mathrm{E}r_0^2 \sum_{s=1}^{\infty} b_{t+s} b_s, \qquad t = 0, 1, \dots. \tag{4.20}$$

Remark 4.2.2 Let $p \ge 2$ and $|\mu|_p < \infty$, then by inequality (4.13), $\{r_t\}$ being a stationary L^p-solution of (4.10) is equivalent to $\{r_t\}$ being a stationary L^2-solution of (4.10) with $\mathrm{E}|r_0|^p < \infty$. Similarly, if Q and $\{\zeta_t\}$ satisfy the conditions of Proposition 4.2.2 and $p \ge 2$, then $\{X_t\}$ being a stationary L^p-solution of (4.12) is equivalent to $\{X_t\}$ being a stationary L^2-solution of (4.12) with $\mathrm{E}|X_0|^p < \infty$.

Proof of Proposition 4.2.2. (i) Since $\{X_t\}$ is predictable and Q satisfies (4.16) so

$$\begin{aligned} \mathrm{E}|r_t|^p &= |\mu|_p \mathrm{E}|Q(a + X_t)|^p \\ &\le |\mu|_p \mathrm{E}|c_1^2 + c_2^2(a + X_t)^2|^{p/2} \\ &\le C(1 + \mathrm{E}|X_t|^p) < C < \infty, \end{aligned}$$

proving (4.17). Moreover, if $p > 1$ then $\mathrm{E}[r_t|\mathcal{F}_{t-1}] = 0$ is a stationary martingale difference sequence. Hence by Proposition 4.2.1, the series in (4.11) converges in L^p and satisfies

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

In particular, $\zeta_t Q(a + \sum_{s<t} b_{t-s} r_s) = \zeta_t Q(a + X_t) = r_t$ by the definition of r_t. Hence, $\{r_t\}$ is a L^p-solution of (4.10). Stationarity of $\{r_t\}$ follows from stationarity of $\{X_t\}$.

Relations (4.18) follow from $\mathrm{E}[\zeta_t|\mathcal{F}_{t-1}] = 0$, $\mathrm{E}[|\zeta_t|^p|\mathcal{F}_{t-1}] = |\mu|_p$, $p > 1$, and the fact that X_t is $\mathcal{F}_{t-1}$-measurable.

(ii) Since $\{r_t\}$ is a L^p-solution of (4.10), so $r_t = \zeta_t Q(a + X_t)$ with X_t defined in (4.11), and $\{X_t\}$ satisfy (4.12), where the series converges in L^p. Also note that $\{X_t\}$

is predictable. Hence, $\{X_t\}$ is a L^p-solution of (4.12). By (4.16),

$$\mathrm{E}|r_t|^p \;=\; |\mu|_p \mathrm{E}|Q(a+X_t)|^p \;\le\; |\mu|_p \mathrm{E}|c_1^2 + c_2^2(a+X_t)^2|^{p/2} \;\le\; C(1+\mathrm{E}|X_t|^p) \;<\; C.$$

It also readily follows that, for $p > 1$, $\{r_t, \mathcal{F}_t, t \in \mathbb{Z}\}$ is a martingale difference sequence. Hence, by the moment inequality in (4.13),

$$\mathrm{E}|X_t|^p \;\le\; K_p \left\{ \begin{array}{ll} \sum_{j=1}^\infty |b_j|^p \mathrm{E}|r_{t-j}|^p, & 0 < p \le 2 \\ \left(\sum_{j=1}^\infty b_j^2 \mathrm{E}^{2/p}|r_{t-j}|^p\right)^{p/2}, & p > 2 \end{array} \right\} \;=\; CB_p \mathrm{E}|r_t|^p, \quad (4.21)$$

proving (4.19). Stationarity of $\{X_t\}$ and (4.20) are easy consequences of the above facts and stationarity of $\{r_t\}$. □

The following theorem obtains a sufficient condition in (4.22) for the existence of a stationary L^p-solution of equations (4.10) and (4.12). Condition (4.22) involves the pth moment of innovations, the Lipschitz constant Lip_Q, the sum B_p in (4.14) and the Rosenthal constant K_p in (4.13). Part (ii) of Theorem 4.2.1 shows that for $p = 2$, condition (4.22) is close to optimal, being necessary in the case of quadratic $Q^2(x) = c_1^2 + c_2^2 x^2$.

Theorem 4.2.1 *Let the conditions of Proposition 4.2.2 be satisfied, $p > 0$ is arbitrary. In addition, assume that Q satisfies the Lipschitz condition in* (4.15)*.*

(i) Let

$$K_p |\mu|_p \mathrm{Lip}_Q^p B_p < 1. \quad (4.22)$$

Then there exists a unique stationary L^p-solution $\{X_t\}$ of (4.12) *and*

$$\mathrm{E}|X_t|^p \;\le \frac{C(p,Q)|\mu|_p B_p}{1 - K_p |\mu|_p \mathrm{Lip}_Q^p B_p}, \quad (4.23)$$

where $C(p,Q) < \infty$ depends only on p and c_1, c_2 in (4.16)*.*

(ii) Assume, in addition, that $Q^2(x) = c_1^2 + c_2^2 x^2$, where $c_i \ge 0, i = 1, 2$, and $\mu_2 = \mathrm{E}\zeta_0^2 = 1$. Then $c_2^2 B_2 < 1$ is a necessary and sufficient condition for the existence of a stationary L^2-solution $\{X_t\}$ of (4.12) *with $a \ne 0$.*

Remark 4.2.3 Condition (4.22) agrees with the contraction condition for the operator defined by the r.h.s. of (4.12) and acting in a suitable space of predictable processes with values in L^p. For the LARCH model, explicit conditions for finiteness of the pth moment were obtained in Giraitis et al. (2000), Giraitis et al. (2004) using a specific diagram approach for multiple Volterra series. For larger values of $p > 2$, condition (4.22) is preferable to the corresponding condition

$$(2^p - p - 1)^{1/2} |\mu|_p^{1/p} B_p^{1/p} < 1, \qquad p = 2, 4, 6, \dots, \tag{4.24}$$

in Giraitis et al. (2000), formula (2.12) for the LARCH model, since the coefficient $(2^p - p - 1)^{1/2}$ grows exponentially with p in contrast to the bound on $K_p^{1/p}$ in Remark 4.2.1 (see also Chapter 3, Example 3.4.3). On the other hand for $p = 4$ (4.24) becomes $\sqrt{11}|\mu|_4^{1/4} B_2^{1/2} < 1$ while (4.22) is satisfied if $K_4^{1/4}|\mu|_4^{1/4} B_2^{1/2} \le 27.083|\mu|_4^{1/4} B_2^{1/2} < 1$, see Remark 4.2.1, which is worse than (4.24).

Proof of Theorem 4.2.1. (i) For $n \in \mathbb{N}$ define a solution of (4.12) with zero initial condition at $t \le -n$ as

$$X_t^{(n)} := \begin{cases} 0, & t \le -n, \\ \sum_{s=-n}^{t-1} b_{t-s} \zeta_s Q(a + X_s^{(n)}), & t > -n, \quad t \in \mathbb{Z}. \end{cases} \tag{4.25}$$

Let us show that $\{X_t^{(n)}\}$ converges in L^p to a stationary L^p-solution $\{X_t\}$ as $n \to \infty$. First, let $0 < p \le 2$. Let $m > n \ge 0$. Then by inequality (4.13) for any $t > -m$ we have that

$$\begin{aligned} \mathrm{E}|X_t^{(m)} - X_t^{(n)}|^p &= K_p |\mu|_p \Big\{ \sum_{-m \le s < -n} |b_{t-s}|^p \mathrm{E}|Q(a + X_s^{(m)})|^p \\ &+ \sum_{-n \le s < t} |b_{t-s}|^p \mathrm{E}\Big|Q(a + X_s^{(n)}) - Q(a + X_s^{(m)})\Big|^p \Big\} \\ &=: K_p |\mu|_p \{S'_{m,n} + S''_{m,n}\}. \end{aligned}$$

Let $\chi_p(n) := \sum_{j=n}^{\infty} |b_j|^p$. From the bound $|a + x|^2 \le (2a^2/\epsilon) + (1 + \epsilon)x^2$, valid for any $0 < \epsilon < 1/2$, $x \in \mathbb{R}$ and $a \ge 0$, it follows that

$$\begin{aligned} \Big|c_1^2 + c_2^2(a + X_s^{(m)})^2\Big|^{p/2} &\le c_1^p + c_2^p |(a + X_s^{(m)})^2|^{p/2} \\ &\le C(c_1, c_2) + c_2^p \, (1 + \epsilon)^{p/2} |X_s^{(m)}|^p \\ &\le C(c_1, c_2) + c_3^p |X_s^{(m)}|^p, \end{aligned}$$

with $c_3 > c_2 > \mathrm{Lip}_Q$ arbitrarily close to Lip_Q. Then using (4.16) we obtain

$$\begin{aligned}
S'_{m,n} &\le \sum_{-m\le s<-n} |b_{t-s}|^p \mathrm{E}\Big|c_1^2 + c_2^2(a+X_s^{(m)})^2\Big|^p \\
&\le C(Q)K_p|\mu|_p\chi_p(t+n) + c_3^p \sum_{-m\le s<-n} |b_{t-s}|^p \mathrm{E}|X_s^{(m)} - X_s^{(n)}|^p, \\
S''_{m,n} &\le \mathrm{Lip}_Q^p \sum_{-n\le s<t} |b_{t-s}|^p \mathrm{E}\Big|X_s^{(n)} - X_s^{(m)}\Big|^p.
\end{aligned}$$

Consequently,

$$\mathrm{E}|X_t^{(m)} - X_t^{(n)}|^p \le C(Q)K_p|\mu|_p\chi_p(t+n) + K_p|\mu|_p c_3^p \sum_{-m\le s<t} |b_{t-s}|^p \mathrm{E}\Big|X_s^{(n)} - X_s^{(m)}\Big|^p.$$

Iterating the above inequality, we obtain

$$\begin{aligned}
\mathrm{E}|X_t^{(m)} - X_t^{(n)}|^p &\le C(Q)K_p|\mu|_p\Big\{\chi_p(t+n) + \sum_{k=1}^{\infty}(K_p|\mu|_k c_3^p)^k \\
&\quad \times \sum_{-m\le s_k<\cdots<s_1<t} |b_{t-s_1}|^p|b_{s_1-s_2}|^p \ldots |b_{s_{k-1}-s_k}|^p \chi_p(s_k+n)\Big\}.
\end{aligned} \tag{4.26}$$

Since $K_p|\mu|_p c_3^p B_p < 1$ by (4.22) and $\sup_{s\ge1}\chi_p(s) \le B_p < \infty$, the series on the r.h.s. of (4.26) is bounded uniformly in m, n and tends to zero as $m, n \to \infty$ by the dominated convergence theorem. Hence, there exist $X_t, t \in \mathbb{Z}$ such that

$$\lim_{n\to\infty} \mathrm{E}|X_t - X_t^{(n)}|^p = 0, \qquad \forall\, t \in \mathbb{Z}. \tag{4.27}$$

Note that $\{X_t\}$ is predictable and

$$\mathrm{E}|X_t|^p = \lim_{n\to\infty} \mathrm{E}|X_t^{(n)}|^p \le \frac{C(Q)K_p|\mu|_p B_p}{1 - K_p|\mu|_p c_3^p B_p} \le \frac{C(p,Q)|\mu|_p B_p}{1 - K_p|\mu|_p \mathrm{Lip}_Q^p B_p},$$

where the last inequality follows by taking $c_3 > \mathrm{Lip}_Q$ sufficiently close to Lip_Q.

We also have by (4.23) and (4.15) that

$$
\begin{aligned}
& \mathrm{E}\Big|\sum_{s<t} b_{t-s}\zeta_s Q(a+X_s) - \sum_{s=-n}^{t-1} b_{t-s}\zeta_s Q(a+X_s^{(n)})\Big|^p \\
= \ & \mathrm{E}\Big|\sum_{s<-n} b_{t-s}\zeta_s Q(a+X_s) + \sum_{s=-n}^{t-1} b_{t-s}\zeta_s (Q(a+X_s) - Q(a+X_s^{(n)}))\Big|^p \\
\le \ & K_p|\mu|_p\Big\{\sum_{s<-n} |b_{t-s}|^p \mathrm{E}\Big|Q(a+X_s)\Big|^p \\
+ \ & \sum_{-n\le s<t} |b_{t-s}|^p \mathrm{E}\Big|Q(a+X_s) - Q(a+X_s^{(n)})\Big|^p\Big\} \\
\le \ & C\Big(\sum_{s<-n}|b_{t-s}|^p + \sum_{s<t}|b_{t-s}|^p\mathrm{E}\Big|X_s - X_s^{(n)}\Big|^p\Big) \ \to \ 0
\end{aligned}
$$

as $n \to \infty$. Whence and from (4.25) it follows that $\{X_t\}$ is a stationary L^p-solution of (4.12) satisfying (4.23).

To show the uniqueness of stationary L^p-solution of (4.12), let $\{X'_t\}, \{X''_t\}$ be two such solutions of (4.12), and $m_p(t) := \mathrm{E}|X'_t - X''_t|^p$. Then $\sup_{t\in\mathbb{Z}} m_p(t) \le M < \infty$ and $m_p(t) \le K_p|\mu|_p \mathrm{Lip}_Q^p \sum_{s<t}|b_{t-s}|^p m_p(s)$ follows by (4.15). Iterating the last equation we obtain that $m_p(t) \le (K_p|\mu|_p\mathrm{Lip}_Q^p B_p)^k M$ holds for all $k \ge 1$, where $K_p|\mu|_p\mathrm{Lip}_Q^p B_p < 1$. Hence, $m_p(t) = 0$. This proves part (i) for $0 < p \le 2$.

The proof of part (i) for $p > 2$ is analogous. Particularly, using (4.13) as in (4.21), we obtain

$$
\begin{aligned}
\mathrm{E}|X_t|^p \ & \le \ K_p|\mu|_p\Big(\sum_{s<t} b_{t-s}^2 \mathrm{E}^{2/p}|Q(a+X_s)|^p\Big)^{p/2} \\
& \le \ K_p|\mu|_p\Big(\sum_{s<t} b_{t-s}^2 (C(Q) + c_3^p\mathrm{E}|X_s|^p)^{2/p}\Big)^{p/2} \\
& \le \ K_p|\mu|_p B_p(C(p,Q) + c_3^p \sup_{s\in\mathbb{Z}} \mathrm{E}|X_s|^p)
\end{aligned}
$$

implying $(1 - K_p|\mu_p|c_3^p B_p)\sup_{t\in\mathbb{Z}}\mathrm{E}|X_t|^p \le C(p,Q)|\mu|_p B_p$ and hence the bound in (4.23) for $p > 2$, by taking c_3 sufficiently close to Lip_Q. This proves part (i).

(ii) Note that $Q(x) = \sqrt{c_1^2 + c_2^2 x^2}$ is a Lipschitz function and satisfies (4.15) with $\mathrm{Lip}_Q = c_2$. Hence by $K_2 = 1$ and part (i), a unique L^2-solution $\{X_t\}$ of (4.12) under the condition $c_2^2 B_2 < 1$ exists. To show the necessity of the last condition, let $\{X_t\}$

be a stationary L^2-solution of (4.12). Then

$$\begin{aligned}
\mathrm{E}X_t^2 &= \sum_{s<t} b_{t-s}^2 \mathrm{E}Q^2(a+X_s) \\
&= \sum_{s<t} b_{t-s}^2 \mathrm{E}\Big(c_1^2 + c_2^2(a+X_s)^2\Big) \\
&= B_2\Big(c_1^2 + c_2^2(a^2 + \mathrm{E}X_t^2)\Big) \;\; > c_2^2 B_2 \mathrm{E}X_t^2
\end{aligned}$$

since $a \neq 0$. Hence, $c_2^2 B_2 < 1$ unless $\mathrm{E}X_t^2 = 0$, or $\{X_t = 0\}$ is a trivial process. Clearly, (4.12) 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. $\square$

Example 4.2.1 (The LARCH model) Let $Q(x) = x$ and $\{\zeta_t\}$ be a standardized i.i.d. sequence with zero mean and unit variance. Then (4.12) becomes the bilinear equation

$$X_t \;=\; \sum_{s<t} b_{t-s}\zeta_s(a+X_s). \tag{4.28}$$

The corresponding conditionally heteroscedastic process $\{r_t = \zeta_t(a+X_t)\}$ in Proposition 4.2.2(i) is the LARCH model discussed in Giraitis et al. (2000), Giraitis et al. (2004) and elsewhere. As shown in Giraitis et al. (2000), Theorem 2.1, equation (4.28) admits a covariance stationary predictable solution if and only if $B_2 = \sum_{j=1}^{\infty} b_j^2 < 1$. Note the last result agrees with Theorem 4.2.1 (ii). A crucial role in the study of the LARCH model is played by the fact that its solution can be written in terms of the convergent orthogonal Volterra series

$$X_t = a\sum_{k=1}^{\infty} \sum_{s_k<\cdots<s_1<t} b_{t-s_1}\ldots b_{s_{k-1}-s_k}\zeta_{s_1}\ldots\zeta_{s_k}.$$

Except for $Q(x) = x$, in other cases of (4.12) including the QARCH model in (4.4), Volterra series expansions are unknown and their usefulness is doubtful.

Example 4.2.2 (Asymmetric ARCH(1)) Consider the model (4.1) with σ_t in (4.5), viz.

$$r_t = \zeta_t\Big(c^2 + (a + br_{t-1})^2\Big)^{1/2}, \tag{4.29}$$

where $\{\zeta_t\}$ are standardized i.i.d. r.v.'s. By Theorem 4.2.1 (ii), equation (4.29) has a unique stationary solution with finite variance $\mathrm{E}r_t^2 = (a^2+c^2)/(1-b^2)$ if and only if $b^2 < 1$.

In parallel, consider the random-coefficient AR(1) equation

$$\widetilde{r}_t = \kappa\varepsilon_t + b\eta_t\widetilde{r}_{t-1}, \tag{4.30}$$

where $\{(\varepsilon_t, \eta_t)\}$ are i.i.d. random vectors with zero mean $\mathrm{E}\varepsilon_t = \mathrm{E}\eta_t = 0$ and unit variances $\mathrm{E}[\varepsilon_t^2] = \mathrm{E}[\eta_t^2] = 1$ and κ, b are real coefficients. As shown in Sentana (1995) (see also Surgailis (2008)), equation (4.30) has a stationary solution with finite variance under the same condition $b^2 < 1$ as (4.29). Moreover, if the coefficients κ and $\rho := \mathrm{E}[\varepsilon_t \eta_t] \in [-1, 1]$ in (4.30) are related to the coefficients a, c in (4.29) as

$$\kappa\rho = a, \qquad \kappa^2 = a^2 + c^2, \tag{4.31}$$

then the processes in (4.29) and (4.30) have the same volatility forms since

$$\begin{aligned} \widetilde{\sigma}_t^2 \; := \; \mathrm{E}[\widetilde{r}_t^2 | \widetilde{r}_s, s < t] \; &= \; \kappa^2 + 2\kappa b\rho + b^2 \widetilde{r}_{t-1}^2 \\ &= \; c^2 + (a + b\widetilde{r}_{t-1})^2 \end{aligned}$$

agrees with the corresponding expression $\sigma_t^2 = c^2 + (a + br_{t-1})^2$ in the case of (4.5).

A natural question is whether the above stationary solutions $\{r_t\}$ and $\{\widetilde{r}_t\}$ of (4.29) and (4.30), with parameters related as in (4.31), have the same (unconditional) finite-dimensional distributions? As shown in (Surgailis (2008), Corollary 2.1), the answer is positive in the case when $\{\zeta_t\}$ and $\{(\varepsilon_t, \eta_t)\}$ are Gaussian sequences. However, the conditionally Gaussian case seems to be the only exception and in general the processes $\{r_t\}$ and $\{\widetilde{r}_t\}$ have different distributions. This can be seen by considering the 3rd conditional moment of (4.29)

$$\mathrm{E}[r_t^3 | r_{t-1}] = \mu_3 \Big(c^2 + (a + br_{t-1})^2 \Big)^{3/2}, \tag{4.32}$$

which is an irrational function of r_{t-1} (unless $\mu_3 = \mathrm{E}\zeta_0^3 = 0$ or $b = 0$), while a similar moment of (4.30)

$$\mathrm{E}[\widetilde{r}_t^3 | \widetilde{r}_{t-1}] = \kappa^3 \nu_{3,0} + 3b\kappa^2 \nu_{2,1} \widetilde{r}_{t-1} + 3b^2 \kappa \nu_{1,2} \widetilde{r}_{t-1}^2 + b^3 \nu_{0,3} \widetilde{r}_{t-1}^3 \tag{4.33}$$

is a cubic polynomial in $\widetilde{r}_{t-1}$, where $\nu_{i,j} := \mathrm{E}[\varepsilon_0^i \eta_0^j]$. Moreover, (4.32) has a constant sign independent of r_{t-1} while the sign of the cubic polynomial in (4.33) changes with $\widetilde{r}_t$ ranging from ∞ to $-\infty$ if the leading coefficient $b^3 \nu_{0,3} \neq 0$.

Using the last observation we can prove that the bivariate distributions of (r_t, r_{t-1}) and $(\widetilde{r}_t, \widetilde{r}_{t-1})$ are different under general conditions on the innovations and the parameters of the two equations. The argument is as follows. Let $b > 0, c > 0, \mu_3 > 0, \nu_{0,3} = \mathrm{E}\eta_0^3 > 0$. Assume that ζ_0 has a bounded strictly positive density function $0 < f(x) < C, x \in \mathbb{R}$ and (ε_0, η_0) has a bounded strictly positive density function $0 < g(x, y) < C, (x, y) \in \mathbb{R}^2$. The above assumptions imply that the distributions of r_t and $\widetilde{r}_t$ have infinite support. Indeed, by (4.29) and the above assumptions we have that $\mathrm{P}(r_t > K) = \int_{\mathbb{R}} \mathrm{P}(c^2 + (a + br_{t-1})^2 > (K/y)^2) f(y)\mathrm{d}y > 0$ for any $K > 0$ since $\lim_{y\to\infty} \mathrm{P}(c^2 + (a + br_{t-1})^2 > (K/y)^2) = 1$. Similarly, $\mathrm{P}(\widetilde{r}_t > K) = \int_{\mathbb{R}^2} \mathrm{P}(\widetilde{r}_{t-1} > (K - \kappa x)/by) g(x, y)\mathrm{d}x\mathrm{d}y > 0$ and $\mathrm{P}(r_t < -K) > 0, \mathrm{P}(\widetilde{r}_t < -K) > 0$ for any $K > 0$. Since $h(x) := \mu_3\Big(c^2 + (a + bx)^2\Big)^{3/2} \ge 1$ for all $|x| > K$ and any sufficiently large $K > 0$, from (4.32) we obtain that for any $K > 0$

$$
\begin{aligned}
\mathrm{E}r_t^3 \mathbf{1}(r_{t-1} > K) &= \mathrm{E}h(r_{t-1})\mathbf{1}(r_{t-1} > K) > 0 \quad \text{and} \\
\mathrm{E}r_t^3 \mathbf{1}(r_{t-1} < -K) &= \mathrm{E}h(r_{t-1})\mathbf{1}(r_{t-1} < -K) > 0.
\end{aligned}
\tag{4.34}
$$

On the other hand, since $\widetilde{h}(x) := \kappa^3\nu_{3,0} + 3b\kappa^2\nu_{2,1}x + 3b^2\kappa\nu_{1,2}x^2 + b^3\nu_{0,3}x^3 \ge 1$ for $x > K$ and $\widetilde{h}(x) \le -1$ for $x < -K$ and K large enough, from (4.33) we obtain that for all sufficiently large $K > 0$

$$
\begin{aligned}
\mathrm{E}\widetilde{r}_t^3 \mathbf{1}(\widetilde{r}_{t-1} > K) &= \mathrm{E}\widetilde{h}(\widetilde{r}_{t-1})\mathbf{1}(\widetilde{r}_{t-1} > K) > 0 \quad \text{and} \\
\mathrm{E}\widetilde{r}_t^3 \mathbf{1}(\widetilde{r}_{t-1} < -K) &= \mathrm{E}\widetilde{h}(\widetilde{r}_{t-1})\mathbf{1}(\widetilde{r}_{t-1} < -K) < 0.
\end{aligned}
\tag{4.35}
$$

Clearly, (4.34) and (4.35) imply that the bivariate distributions of (r_t, r_{t-1}) and $(\widetilde{r}_t, \widetilde{r}_{t-1})$ are different under the stated assumptions.

For models (4.29) and (4.30), we can explicitly compute covariances $\rho(t) = \mathrm{Cov}(r_t^2, r_0^2)$, $\widetilde{\rho}(t) = \mathrm{Cov}(\widetilde{r}_t^2, \widetilde{r}_0^2)$ and some other joint moment functions, as follows.

Let $\mu_3 = \mathrm{E}\zeta_0^3 = 0, \mu_4 = \mathrm{E}\zeta_0^4 < \infty$ and $m_2 := \mathrm{E}r_0^2$, $m_3(t) := \mathrm{E}r_t^2 r_0$, $m_4(t) := \mathrm{E}r_t^2 r_0^2, t \ge 0$. Then

$$
\begin{aligned}
m_2 &= (a^2 + c^2)/(1 - b^2), \qquad m_3(0) \;=\; 0, \\
m_3(1) &= \mathrm{E}[((a^2 + c^2) + 2abr_0 + b^2 r_0^2) r_0] = 2abm_2 + b^2 m_3(0) \;=\; 2abm_2, \\
m_3(t) &= \mathrm{E}[((a^2 + c^2) + 2abr_{t-1} + b^2 r_{t-1}^2) r_0] \;=\; b^2 m_3(t-1) \;=\; \ldots \;=\; b^{2(t-1)} m_3(1) \\
&= \frac{2ab(a^2 + c^2)}{1 - b^2} b^{2(t-1)}, \quad t \ge 1.
\end{aligned}
\tag{4.36}
$$

Similarly,

$$
\begin{aligned}
m_4(0) &= \mu_4 \mathrm{E}[((a^2 + c^2) + 2abr_{-1} + b^2 r_{-1}^2)^2] \\
&= \mu_4\{(a^2 + c^2)^2 + (2ab)^2 m_2 + b^4 m_4(0) + 2b^2(a^2 + c^2)m_2\}, \\
m_4(t) &= \mathrm{E}[((a^2 + c^2) + 2abr_{t-1} + b^2 r_{t-1}^2) r_0^2] = (a^2 + c^2)m_2 + b^2 m_4(t-1), \quad t \ge 1
\end{aligned}
$$

resulting in

$$
\begin{aligned}
m_4(0) &= \frac{\mu_4((a^2+c^2)^2 + ((2ab)^2 + 2(a^2+c^2)b^2)m_2)}{1-\mu_4 b^4}, \\
m_4(t) &= m_2(a^2+c^2)\cdot\frac{1-b^{2t}}{1-b^2} + b^{2t}m_4(0), \quad t\ge 1,
\end{aligned}
\tag{4.37}
$$

and

$$
\rho(t) = (m_4(0)-m_2^2)b^{2t}, \qquad t\ge 0. \tag{4.38}
$$

In a similar way, when the distribution of ζ_0 is symmetric one can write recursive linear equations for joint even moments $\mathrm{E}[r^{2p}(0)r^{2p}(t)]$ of arbitrary order $p=1,2,\dots$ involving $\mathrm{E}[r^{2l}(0)r^{2p}(t)], 1\le l\le p-1$ and $m_{2k}(0)=\mathrm{E}[r^{2k}(0)], 1\le k\le 2p$. These equations can be explicitly solved in terms of a,b,c and $\mu_{2k}, 1\le k\le 2p$.

A similar approach can be applied to find joint moments of the random-coefficient AR(1) process in (4.30), with the difference that symmetry of (ε_0,η_0) is not needed. Let $\widetilde{m}_2 := \mathrm{E}\tilde{r}_t^2$, $\widetilde{m}_3(t) := \mathrm{E}[\tilde{r}_t^2\tilde{r}_0]$, $\widetilde{m}_4(t) := \mathrm{E}[\tilde{r}_t^2\tilde{r}_0^2]$ and $\tilde{\rho}(t) := \mathrm{Cov}(\tilde{r}_t^2,\tilde{r}_0^2)$, $\nu_{i,i} := \mathrm{E}[\varepsilon_0^i\eta_0^j]$. Then

$$
\begin{aligned}
\widetilde{m}_2 &= \kappa^2/(1-b^2), \\
\widetilde{m}_3(0) &= \mathrm{E}[(\kappa\varepsilon_0 + b\eta_0\tilde{r}_{-1})^3] \;=\; \kappa^3\nu_{3,0} + 3\kappa b^2\nu_{1,2}\widetilde{m}_2 + b^3\nu_{0,3}\widetilde{m}_3(0), \\
\widetilde{m}_3(1) &= \mathrm{E}[(\kappa + 2\kappa\rho b\tilde{r}_0 + b^2\tilde{r}_0^2)\tilde{r}_0] \;=\; 2\kappa\rho b\widetilde{m}_2 + b^2\widetilde{m}_3(0), \\
\widetilde{m}_3(t) &= \mathrm{E}[(\kappa^2 + 2\kappa\rho b\tilde{r}_{t-1} + b^2\tilde{r}_{t-1}^2)\tilde{r}_0] \\
&= b^2\widetilde{m}_3(t-1) \;=\cdots=\; b^{2(t-1)}\widetilde{m}_3(1), \quad t\ge 2
\end{aligned}
$$

and

$$
\begin{aligned}
\widetilde{m}_4(0) &= \mathrm{E}[(\kappa\varepsilon_0 + b\eta_0\tilde{r}_{-1})^4] \\
&= \kappa^4\nu_{4,0} + 6\kappa^2 b^2\nu_{2,2}\widetilde{m}_2 + 4\kappa b^3\nu_{1,3}\widetilde{m}_3(0) + b^4\nu_{0,4}\widetilde{m}_4(0), \\
\widetilde{m}_4(1) &= \mathrm{E}[(\kappa\varepsilon_t + b\eta_t\tilde{r}_0)^2\tilde{r}_0^2] = \kappa^2\widetilde{m}_2 + 2\kappa\rho b\widetilde{m}_3(0) + b^2\widetilde{m}_4(0), \\
\widetilde{m}_4(t) &= \mathrm{E}[(\kappa\varepsilon_t + b\eta_t\tilde{r}_{t-1})^2\tilde{r}_0^2] \;=\; \kappa^2\widetilde{m}_2 + b^2\widetilde{m}_4(t-1), \qquad t\ge 2,
\end{aligned}
$$

leading to

$$\begin{aligned}
\widetilde{m}_3(0) &= \frac{\kappa^3\nu_{3,0} + 3\kappa b^2\nu_{1,2}\widetilde{m}_2}{1-\nu_{0,3}b^3}, \\
\widetilde{m}_3(t) &= b^{2(t-1)}(2\kappa\rho b\widetilde{m}_2 + b^2\widetilde{m}_3(0)), \quad t\geq 1, \\
\widetilde{m}_4(0) &= \frac{\kappa^4\nu_{4,0} + 6\kappa^2b^2\nu_{2,2}\widetilde{m}_2 + 4\kappa b^3\nu_{1,3}\widetilde{m}_3(0)}{1-\nu_{0,4}b^4}, \\
\widetilde{m}_4(t) &= \widetilde{m}_2\kappa^2\Big(\frac{1-b^{2t}}{1-b^2}\Big) + b^{2t}(\widetilde{m}_4(0) + 2\kappa\rho\widetilde{m}_3(0)/b), \quad t\geq 1,
\end{aligned} \tag{4.39}$$

and

$$\begin{aligned}
\widetilde{\rho}_4(t) &= b^{2(t-1)}\widetilde{\rho}_4(1), \quad t\geq 1, \\
\widetilde{\rho}_4(1) &= 2\rho\kappa b\widetilde{m}_3(0) + b^2(\widetilde{m}_4(0) - \widetilde{m}_2^2).
\end{aligned}$$

Then if $\nu_{3,0} = \nu_{1,2} = 0$ we have $\widetilde{m}_3(0) = 0$ and $\widetilde{\rho}_4(t) = (\widetilde{m}_4(0) - \widetilde{m}_2^2)b^{2t}$; moreover, $\widetilde{m}_2 = m_2$ in view of (4.31). Then $\widetilde{\rho}_4(t) = \rho_4(t)$ is equivalent to $\widetilde{m}_4(0) = m_4(0)$, which follows from

$$\mu_4 = \nu_{0,4} = \nu_{4,0} \quad \text{and} \quad 6\nu_{2,2} = \mu_4(4\nu_{1,1}^2 + 2), \tag{4.40}$$

see (4.38), (4.37), (4.39). Note that (4.40) hold for centered Gaussian distribution (ε_0, η_0) with unit variances $\mathrm{E}\varepsilon_0^2 = \mathrm{E}\eta_0^2 = 1$.

4.3 Weak dependence

Various measures of weak dependence for stationary processes $\{y_t\} = \{y_t, t\in\mathbb{Z}\}$ have been introduced in the literature, see e.g. Dedecker et al. (2007). Usually, the dependence between the present ($t\geq 0$) and the past ($t\leq -n$) values of $\{y_t\}$ is measured by some dependence coefficients decaying to 0 as $n\to\infty$. The decay rate of these coefficients plays a crucial role in establishing many asymptotic results. The classical problem is proving Donsker's invariance principle:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}(y - \mathrm{E}y_t) \to \sigma B(\tau), \qquad \text{in the Skorohod space } \ D[0,1], \tag{4.41}$$

where $B = \{B(\tau), \tau\in[0,1]\}$ is a standard Brownian motion. The above result is useful in change-point analysis (Csörgő and Horváth (1997)), financial mathematics and many other areas. Further applications of weak dependence coefficients include empirical processes in Dedecker and Prieur (2007) and the asymptotic behavior of various statistics, including the maximum likelihood estimators (see Ibragimov and Linnik (1971) and the application to GARCH estimation in Lindner (2009)).

The present section discusses two measures of weak dependence - the projective weak dependence coefficients of Wu (2005) and the τ-dependence coefficients introduced in Dedecker and Prieur (2004), Dedecker and Prieur (2005) - for stationary solutions $\{r_t\}, \{X_t\}$ of equations (4.10), (4.12). We show that the decay rate of the above weak dependence coefficients is determined by the decay rate of the moving average coefficients b_j.

Projective weak dependence coefficients

Let us introduce some notation. For r.v. ξ, write $\|\xi\|_p := \mathrm{E}^{1/p}|\xi|^p$, $p \geq 1$. Let $\{y_t, t \in \mathbb{Z}\}$ be a stationary causal Bernoulli shift in i.i.d. sequence $\{\zeta_t\}$, in other words,

$$y_t = f(\zeta_s, s \leq t), \quad t \in \mathbb{Z}, \tag{4.42}$$

where $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ is a measurable function. We also assume $\mathrm{E}y_0 = 0, \|y_0\|_p = \mathrm{E}^{1/p}|y_0|^p < \infty$. Introduce the projective weak dependence coefficients

$$\omega_p(i;\{y_t\}) \ := \ \|f_i(\xi_0) - f_i(\xi_0')\|_p, \qquad \delta_p(i;\{y_t\}) \ := \ \|f(\xi_i) - f(\xi_i')\|_p, \tag{4.43}$$

where $\xi_i := (\ldots, \zeta_{-1}, \zeta_0, \zeta_1, \ldots, \zeta_i), \xi_i' := (\ldots, \zeta_{-1}, \zeta_0', \zeta_1, \ldots, \zeta_i), f_i(\xi_0) := \mathrm{E}[f(\xi_i)|\xi_0] = \mathrm{E}[y_i|\mathcal{F}_0]$ is the conditional expectation and $\{\zeta_0', \zeta_t, t \in \mathbb{Z}\}$ are i.i.d. r.v.s. Note the i.i.d. sequences ξ and ξ_i' coincide except for a single entry. Then $\omega_p(i;\{y_t\}) \leq \delta_p(i;\{y_t\}), i \geq 0$ and condition

$$\sum_{k=0}^{\infty} \omega_2(k;\{y_t\}) \ < \infty \tag{4.44}$$

guarantees the weak invariance principle in (4.41) with $\sigma^2 := \sum_{j\in\mathbb{Z}} \mathrm{Cov}(y_0, y_j)$, see Wu (2005). The last series absolutely converges in view of (4.44) and the bound in Wu (2005), Theorem 1, implying $|\mathrm{Cov}(y_0, y_j)| \leq \sum_{k=0}^{\infty} \omega_2(k;\{y_t\})\omega_2(k+j;\{y_t\}), j \geq 0$.

Below, we verify Wu's condition (4.44) for $\{X_t\}, \{r_t\}$ in (4.12), (4.10). We assume that the coefficients b_j decay as $j^{-\gamma}$ with some $\gamma > 1$, viz.,

$$\exists \gamma > 0, \ c > 0 : \qquad |b_j| < cj^{-\gamma}, \quad \forall \ j \geq 1. \tag{4.45}$$

Proposition 4.3.1 *Let Q be Lipschitz function as in* (4.15)*, $p \geq 1$, $K_p|\mu|_p \mathrm{Lip}_Q^p B_p < 1$ (see* (4.22)*), and $\{X_t\}, \{r_t\}$ be stationary L^p-solutions of* (4.12), (4.10)*, respectively.*

In addition, assume that b_j *satisfy* (4.45) *with* $\gamma > \max\{1/2, 1/p\}$.
Then

$$\delta_p(k; \{X_t\}) = O(k^{-\gamma}) \qquad \textit{and} \qquad \delta_p(k; \{r_t\}) = O(k^{-\gamma}). \tag{4.46}$$

Proof. We will give the proof for $p \ge 2$ only as the proof for $p \in [1,2]$ is similar.

Following the notation in (4.43), let $\{X'_t\}, \{r'_t\}$ be the corresponding processes (Bernoulli shifts) of the i.i.d. sequence $\xi' := (\dots, \zeta_{-1}, \zeta'_0, \zeta_1, \zeta_2, \dots)$ with ζ_0 replaced by its independent copy ζ'_0. Note that $X'_t = X_t \, (t \le 0)$, $r'_t = r_t \, (t < 0)$. We have $\delta_2^2(k; \{X_t\}) = (\mathrm{E}|X_k - X'_k|^p)^{2/p} = \|X_k - X'_k\|_p^2$, where

$$X_k - X'_k \;=\; \sum_{0<s<k} b_{t-s}(r_s - r'_s) + b_k(\zeta_0 - \zeta'_0)Q(a + X_0).$$

Then with $v_p^2 := \|Q(a+X_0)\|_p^2$ using Rosenthal's inequality (4.13) similarly as in the proof of Theorem 4.2.1 we obtain

$$\begin{aligned}
\|X_k - X'_k\|_p^2 \;&\le\; K_p^{2/p}\Big(\sum_{0<s<k} b_{k-s}^2 \|r_s - r'_s\|_p^2 + \|\zeta_0 - \zeta'_0\|_p^2 b_k^2 v_p^2\Big) \\
&\le\; K_p^{2/p}\Big(\sum_{0<s<k} b_{k-s}^2 |\mu|_p^{2/p} \|Q(a+X_s) - Q(a+X'_s)\|_p^2 + 4|\mu|_p^{2/p} b_k^2 v_p^2\Big) \\
&\le\; K_p^{2/p}|\mu|_p^{2/p}\Big(\mathrm{Lip}_Q^2 \sum_{0<s<k} b_{k-s}^2 \|X_s - X'_s\|_p^2 + 4 b_k^2 v_p^2\Big).
\end{aligned}$$

Let $\alpha_k := K_p^{2/p}|\mu|_p^{2/p}\mathrm{Lip}_Q^2 b_k^2$. Iterating the last inequality we obtain

$$\delta_2^2(k; \{X_t\}) \;\le\; \frac{4v_p^2}{\mathrm{Lip}_Q^2}\Big(\alpha_k + \sum_{0<s<k} \alpha_s \alpha_{k-s} + \dots\Big) \;=\; \frac{4v_p^2}{\mathrm{Lip}_Q^2} \cdot A_k,$$

where A_k is as in (A.1) (see Appendix A). Since $A = \sum_{k>0} \alpha_k = (K_p|\mu|_p \mathrm{Lip}_Q^p B_p)^{2/p} < 1$ and $\alpha_k \le Ck^{-2\gamma}$, by Lemma A.1 we obtain $\delta_2(k; \{X_t\}) \le Ck^{-\gamma}$, proving the first inequality in (4.46). The proof of the second inequality in (4.46) follows similarly using $\delta_p^2(k; \{r_t\}) = \|r_k - r'_k\|_p^2 \le \mathrm{Lip}_Q^2 |\mu|_p^{2/p} \|X_k - X'_k\|_p^2 = \mathrm{Lip}_Q^2 |\mu|_p^{2/p} \delta_2^2(k; \{X_t\})$. $\square$

The next corollary follows from the above-mentioned result of Wu (2005), relations $\delta_2(k; \{y_t\}) \le C\delta_2(k; \{r_t\})$, $\delta_2(k; \{z_t\}) \le C\delta_2(k; \{X_t\})$ and (4.46).

Corollary 4.3.1 *Let* $\{y_t := h(r_t)\}, \{z_t := h(X_t)\}$*, where* $\{X_t\}, \{r_t\}$ *are as in Proposition 4.3.1,* $p = 2$ *and* $h : \mathbb{R} \to \mathbb{R}$ *is a Lipschitz function. In addition, assume that* b_j *satisfy* (4.45) *with* $\gamma > 1$*. Then*

$$n^{-1/2} \sum_{t=1}^{[n\tau]} (y_t - \mathrm{E}y_t) \;\to_{D[0,1]}\; c_y B(\tau) \quad \textit{and} \quad n^{-1/2} \sum_{t=1}^{[n\tau]} (z_t - \mathrm{E}z_t) \;\to_{D[0,1]}\; c_z B(\tau), \tag{4.47}$$

where B is a standard Brownian motion and

$$c_y^2 := \sum_{t\in\mathbb{Z}} \mathrm{Cov}(y_0, y_t) < \infty, \qquad c_z^2 := \sum_{t\in\mathbb{Z}} \mathrm{Cov}(z_0, z_t) < \infty.$$

τ-weak dependence coefficients

Let $\{y_t, t \in \mathbb{Z}\}$ be a stationary process with $\|y_0\|_p < \infty, p \in [1,\infty]$. Following Dedecker and Prieur (2004), Dedecker and Prieur (2005), we define the τ-weak dependence coefficients

$$\tau_p(\{y_{j_i}\}_{1\le i\le k}) \ := \ \Big\| \sup_{f\in\Lambda_1(\mathbb{R}^k)} \Big| \mathrm{E}\Big[f(y_{j_1},\dots,y_{j_k})\Big| y_t, t\le 0\Big] - \mathrm{E}[f(y_{j_1},\dots,y_{j_k})]\Big|\Big\|_p$$

measuring the dependence between $\{y_t\}_{t\le 0}$ and $\{y_{j_i}\}_{1\le i\le k}, 0 < j_1 < \cdots < j_k$, and

$$\tau_p(n; \{y_j\}) \ := \ \sup_{k\ge 1} k^{-1} \sup_{n\le j_1<\cdots<j_k} \tau_p(\{y_{j_i}\}_{1\le i\le k}).$$

Here, $\Lambda_1(\mathbb{R}^k)$ denotes the class of all Lipschitz functions $f : \mathbb{R}^k \to \mathbb{R}$ with

$$|f(x_1,\dots,x_k) - f(y_1,\dots,y_k)| \le \sum_{i=1}^k |x_i - y_i| \quad \text{for any} \quad (x_1,\dots,x_k),\ (y_1,\dots,y_k) \in \mathbb{R}^k.$$

In the case when $\{y_t\}$ is a causal Bernoulli shift of an i.i.d. sequence $\{\zeta_t\}$ as in (4.42), τ-coefficients can be estimated via δ-coefficients:

$$\tau_p(n, \{y_t\}) \ \le \ \sum_{j=n}^\infty \delta_p(j, \{y_t\}). \tag{4.48}$$

The above bound is an easy consequence of the coupling inequality of Dedecker and Prieur (2005):

$$\tau_p(\{y_{j_i}\}_{1\le i\le k}) \ \le \ \sum_{i=1}^k \|y_{j_i} - y^*_{j_i}\|_p \quad \text{and} \quad \tau_p(n, \{y_t\}) \ \le \ \sup_{j\ge n} \|y_j - y^*_j\|_p, \tag{4.49}$$

where $\{y^*_j\}$ has the same distribution as $\{y_t\}$ and is independent of $y_s, s \le 0$. Indeed, let $\{y^*_t\}$ be the corresponding process (Bernoulli shift) of the i.i.d. sequence $\xi^* := (\dots,\zeta^*_{-2},\zeta^*_{-1},\zeta_0,\zeta_1,\dots)$ with $(\zeta^*_s, s<0)$ an independent copy of $(\zeta_s, s<0)$. Introduce also "intermediate" i.i.d. sequence $\xi^*_i := (\dots,\zeta^*_{-i-1},\zeta^*_{-i},\zeta_{-i+1},\dots,\zeta_0,\ \zeta_1,\dots)$, $i \ge 1$, $\xi^*_1 := \xi^*$ and the corresponding Bernoulli shift $\{y^*_{i,t}\}$ with the same f as in (4.42). Note the sequences ξ^*_i and ξ^*_{i+1} agree up to single entry. By triangle inequality, $\|y_n - y^*_n\|_p \le \sum_{i=1}^\infty \|y^*_{i,n} - y^*_{i+1,n}\|_p = \sum_{i=1}^\infty \delta_p(n+i, \{y_t\})$, leading to (4.48) via (4.49).

The following corollary is immediate from (4.48) and Proposition 4.3.1.

Corollary 4.3.2 *Let Q be Lipschitz function as in* (4.15)*, $p \ge 1$, $K_p|\mu|_p \mathrm{Lip}_Q^p B_p < 1$ (see* (4.22)*) and let $\{X_t\}, \{r_t\}$ be stationary L^p-solutions of* (4.12)*,* (4.10)*, respectively. In addition, assume that b_j satisfy* (4.45) *with $\gamma > 1$. Then*

$$\tau_p(n; \{X_j\}) = O(n^{-\gamma+1}), \qquad \tau_p(n; \{r_j\}) = O(n^{-\gamma+1}). \tag{4.50}$$

Theorem 1 in Dedecker and Prieur (2007) together with Corollary 4.3.2 imply the following CLT for the empirical distribution functions $F_n^X(u) := n^{-1}\sum_{t=1}^n \mathbf{1}(X_t \le u)$, $F_n^r(u) := n^{-1}\sum_{t=1}^n \mathbf{1}(r_t \le u)$, $u \in \mathbb{R}$ of stationary solutions $\{X_t\}, \{r_t\}$ of (4.12), (4.10). Let $F^X(u) := \mathrm{P}(X_0 \le u), F^r(u) := \mathrm{P}(r_0 \le u)$ be the corresponding distribution functions. See Dedecker and Prieur (2007) for the definition of weak convergence in the space $\ell^\infty(\mathbb{R})$ of all bounded functions on $\mathbb{R}$.

Corollary 4.3.3 *Let the conditions of Corollary 4.3.2 hold with $p = 1$ and $\gamma > 5$. Moreover, assume that F^X, F^r have bounded densities. Then $\{\sqrt{n}(F_n^X(u) - F^X(u)), u \in \mathbb{R}\}$ and $\{\sqrt{n}(F_n^r(u) - F^r(u)), u \in \mathbb{R}\}$ converge weakly in $\ell^\infty(\mathbb{R})$ as $n \to \infty$ towards Gaussian processes on $\mathbb{R}$ with zero mean and respective covariance functions*

$$\sum_{k\in\mathbb{Z}} \mathrm{Cov}(\mathbf{1}(X_0 \le u), \mathbf{1}(X_k \le u)) \quad \textit{and} \quad \sum_{k\in\mathbb{Z}} \mathrm{Cov}(\mathbf{1}(r_0 \le u), \mathbf{1}(r_k \le u)).$$

Remark 4.3.1 Let the noise ζ_0 have a bounded density f_ζ and $\inf_{x\in\mathbb{R}} Q(x) > 0$, then F^X, F^r have bounded densities f_X, f_r, particularly, $f_r(x) = \mathrm{E}[f_\zeta(x/\sigma_0)/\sigma_0], \sigma_0 = Q(a + X_0)$. Thus, the requirement that F^X, F^r have bounded densities does not necessarily impose additional conditions on the coefficients b_j.

Remark 4.3.2 In Dedecker and Prieur (2007) a general tightness condition is proposed in Proposition 6 for alternative classes of functions besides indicators of half-lines. Conditions are not immediate to check which explains why we restricted to the case of empirical cumulative distribution functions.

4.4 Strong dependence

The term *strong dependence* or *long memory* usually refers to stationary process $\{y_t, t \in \mathbb{Z}\}$ whose covariance decays slowly with the lag so that its absolute series diverges: $\sum_{k=1}^\infty |\mathrm{Cov}(y_0, y_k)| = \infty$. Since the variance of $\sum_{k=1}^n y_t$ usually grows faster than n under long memory, Donsker's invariance principle in (4.41) is no more valid and the limit of the partial sums process, if exists, might be quite complicated.

It is natural to expect that the "long memory" asymptotics of b_j in (4.51) induces some kind of long memory of solutions $\{r_t\}, \{X_t\}$ of (4.1), (4.12), under general assumptions on Q. Concerning the latter process, this is true indeed as shown in the following theorem.

Theorem 4.4.1 *Let $\{X_t\}$ be a stationary L^2-solution of* (4.12)*, where*

$$b_j \ \sim \ \beta\, j^{d-1} \qquad (\exists\ 0 < d < 1/2,\ \beta > 0) \tag{4.51}$$

and Q satisfies the Lipschitz condition in (4.15) *with* $\mathrm{Lip}_Q^2 B^2 = \mathrm{Lip}_Q^2 \sum_{j=1}^{\infty} b_j^2 < 1$. *Then*

$$\begin{aligned} \mathrm{Cov}(X_0, X_t) \ &\sim \ \lambda_1^2 t^{2d-1}, \quad t \to \infty \qquad \textit{and} \\ n^{-d-(1/2)} \sum_{t=1}^{[n\tau]} X_t \ &\to_{D[0,1]} \ \lambda_2 B_{d+(1/2)}(\tau), \end{aligned} \tag{4.52}$$

where $B_{d+(1/2)}$ is a fractional Brownian motion with $\mathrm{Var}(B_{d+(1/2)}(\tau)) = \tau^{2d+1}$ *and* $\lambda_1^2 := \beta^2 B(d, 1-2d) \mathrm{E} Q^2(a + X_0)$, $\lambda_2^2 := \lambda_1^2 / d(1+2d)$.

Proof. The first relation in (4.52) follows from (4.20) and (4.51). The second relation in (4.52) follows from a general result in Abadir et al. (2014), Proposition 3.1, using the fact that $\{r_s\}$ in (4.12) is a stationary ergodic martingale difference sequence. □

Clearly, properties as in (4.52) do not hold for $\{r_t = \zeta_t Q(a + X_t)\}$ which is an uncorrelated martingale difference sequence. Here, long memory should appear in the behavior of the volatility $\sigma_t = Q(a + X_t)$, being "hidden" inside of nonlinear kernel Q. The last fact makes it much harder to prove it rigorously. Further we restrict ourselves to the quadratic model with $Q^2(x) = c^2 + x^2$, or

$$r_t \ = \ \zeta_t \sqrt{c^2 + \Big(a + \sum_{s<t} b_{t-s} r_s\Big)^2}, \qquad t \in \mathbb{Z} \tag{4.53}$$

as in (4.4), where (recall) $\{\zeta_t\}$ are standardized i.i.d. r.v.s, with zero mean and unit variance, and $b_j, j \ge 1$ are real numbers satisfying (4.51).

The following theorem shows that under some additional conditions the squared process $\{r_t^2\}$ of (4.53) has similar long memory properties as $\{X_t\}$ in Theorem 4.4.1. For the LARCH model (see Example 4.2.1 above) similar results were obtained in Giraitis et al. (2000), Theorems 2.2, 2.3.

Theorem 4.4.2 *Let $\{r_t\}$ be a stationary L^2-solution of* (4.53) *with b_j satisfying* (4.51) *and $B^2 = \sum_{j=1}^{\infty} b_j^2 < 1$. Assume in addition that $\mu_4 = \mathrm{E}[\zeta_0^4] < \infty$ and $\mathrm{E}r_t^4 < \infty$. Then*

$$\mathrm{Cov}(r_0^2, r_t^2) \ \sim \ \kappa_1^2 t^{2d-1}, \qquad t \to \infty, \tag{4.54}$$

where $\kappa_1^2 := \left(\frac{2a\beta}{1-B^2}\right)^2 B(d, 1-2d)\mathrm{E}r_0^2$. Moreover,

$$n^{-d-1/2} \sum_{t=1}^{[n\tau]} (r_t^2 - \mathrm{E}r_t^2) \ \to_{D[0,1]} \ \kappa_2 B_{d+1/2}(\tau), \qquad n \to \infty, \tag{4.55}$$

where $B_{d+(1/2)}$ is a fractional Brownian motion as in (4.52) *and $\kappa_2^2 := \kappa_1^2/(d(1+2d))$.*

Proof. The proof of Theorem 4.4.2 heavily relies on the decomposition

$$(r_t^2 - \mathrm{E}r_t^2) - \sum_{s<t} b_{t-s}^2 (r_s^2 - \mathrm{E}r_s^2) \ = \ 2aX_t + Z_t, \tag{4.56}$$

where $\{Z_t\}$ on the r.h.s. of (4.56) is negligible so its memory intensity is less than the memory intensity of the main term, $\{X_t\}$. Accordingly, $r_t^2 - \mathrm{E}r_t^2 = (1 - \sum_{j=1}^{\infty} b_j^2 L^j)^{-1} \xi_t$ behaves like an AR(∞) process with long memory innovations $\xi_t := 2aX_t + Z_t \approx 2aX_t$. A rigorous meaning to the above heuristic explanation is provided below.

By the definition of r_t in (4.53),

$$\begin{aligned} Z_t \ &:= \ U_t + V_t, \qquad \text{where} \\ U_t \ &:= \ (\zeta_t^2 - 1)Q^2(a + X_t), \\ V_t \ &:= \ X_t^2 - \mathrm{E}X_t^2 - \sum_{s<t} b_{t-s}^2 (r_s^2 - \mathrm{E}r_s^2) \\ &= \ 2 \sum_{s_2 < s_1 < t} b_{t-s_1} b_{t-s_2} r_{s_1} r_{s_2}. \end{aligned} \tag{4.57}$$

Let us first check that the double series in (4.57) converges in mean square and (4.57) holds. Let

$$X_{t,N} := \sum_{-N<s<t} b_{t-s} r_s, \qquad V_{t,N} := 2 \sum_{-N<s_2<s_1<t} b_{t-s_1} b_{t-s_2} r_{s_1} r_{s_2},$$

then $V_{t,N} = X_{t,N}^2 - \mathrm{E}X_{t,N}^2 - \sum_{-N<s<t} b_{t-s}^2 (r_s^2 - \mathrm{E}r_s^2)$ and, for $M > N$,

$$\mathrm{E}(X_{t,N}^2 - X_{t,M}^2)^2 = \mathrm{E}(X_{t,N} - X_{t,M})^2 (X_{t,N} + X_{t,M})^2 \le \|X_{t,N} - X_{t,M}\|_4^2 \|X_{t,N} + X_{t,M}\|_4^2.$$

By Rosenthal's inequality in (4.13),

$$\begin{aligned}
\|X_{t,N} + X_{t,M}\|_4^2 &\le C \sum_{-M<s<t} b_{t-s}^2 \le C \qquad \text{and} \\
\|X_{t,N} - X_{t,M}\|_4^2 &\le C \sum_{-M<s\le -N} b_{t-s}^2 \to 0 \qquad (N, M \to \infty).
\end{aligned}$$

Therefore, $\lim_{N,M\to\infty} \mathrm{E}(X_{t,N}^2 - X_{t,M}^2)^2 = 0$.

The convergence of $\mathrm{E}X_{t,N}^2$ and $\sum_{-N<s<t} b_{t-s}^2(r_s^2 - \mathrm{E}r_s^2)$ in L^2 as $N \to \infty$ is easy. Hence, $V_{t,N}, N \ge 1$ is a Cauchy sequence in L^2 and the double series in (4.57) converges as claimed above, proving (4.57).

Let us prove that in the decomposition (4.56), $\{Z_t\}$ is negligible in the sense that its (cross)covariances decay faster as the covariance of the main term, $\{X_t\}$, viz.,

$$\mathrm{E}[Z_t Z_0] = o(t^{2d-1}), \qquad \mathrm{E}[X_t Z_0] = o(t^{2d-1}), \qquad \mathrm{E}[Z_t X_0] = o(t^{2d-1}) \tag{4.58}$$

as $t \to \infty$. Note, for $t \ge 1$, $\mathrm{E}[U_0 U_t] = \mathrm{E}[V_0 U_t] = 0$ and $\mathrm{E}[V_t U_0] = 2b_t \mathrm{E}[\zeta_0(\zeta_0^2 - 1)Q^2(a + X_0)\sum_{s_2<0} b_{t-s_2} r_{s_2}] = O(b_t) = o(t^{2d-1})$. Hence, the first relation in (4.58) follows from

$$\mathrm{E}[V_t V_0] = o(t^{2d-1}), \qquad t \to \infty, \tag{4.59}$$

which is proved below. Since $\mathrm{E}[V_t^2] < \infty, \mathrm{E}[V_t] = 0$ we can write the orthogonal expansion

$$V_t = \sum_{s<t} P_s V_t,$$

where $P_s V_t := \mathrm{E}[V_t|\mathcal{F}_s] - \mathrm{E}[V_t|\mathcal{F}_{s-1}]$ is the projection operator.

By orthogonality of P_s,

$$\Big|\mathrm{E}V_0 V_t\Big| = \Big|\sum_{s<0} \mathrm{E}[(P_s V_0)(P_s V_t)]\Big| \le \sum_{s<0} \|P_s V_0\|_2 \|P_s V_t\|_2.$$

Relation (4.59) follows from

$$\|P_s V_0\|_2^2 = o(b_{-s}^2) = o((-s)^{2(d-1)}), \qquad s \to -\infty. \tag{4.60}$$

Indeed, if (4.60) is true then

$$\mathrm{E}V_0 V_t = o\Big(\sum_{s<0} (-s)^{d-1}(t-s)^{d-1}\Big) = o(t^{2d-1}), \qquad t \to \infty,$$

proving (4.59).

Consider (4.60). We have by (4.57) and the martingale difference property of $\{r_s\}$

that

$$P_s V_0 \;=\; 2r_s b_{-s} \sum_{u<s} b_{-u} r_u$$

and

$$\|P_s V_0\|_2^2 \;=\; 4b_{-s}^2 \mathrm{E}\Big[r_s^2\Big(\sum_{u<s} b_{-u} r_u\Big)^2\Big] \;\le\; 4b_{-s}^2 \|r_s\|_4^2 \Big\|\sum_{u<s} b_{-u} r_u\Big\|_4^2.$$

By Rosenthal's inequality in (4.13),

$$\mathrm{E}\Big|\sum_{u<s} b_{-u} r_u\Big|^4 \;\le\; C_4\Big(\sum_{u<s} b_{-u}^2 (\mathrm{E}r_u^4)^{1/2}\Big)^2 \le C\Big(\sum_{u>|s|} u^{2(d-1)}\Big)^2 = O(|s|^{2(2d-1)}) = o(1).$$

Therefore,

$$\|P_s V_0\|_2^2 \;\le\; C|s|^{2(d-1)+2d-1} \;=\; o(|s|^{2(d-1)}),$$

proving (4.60), (4.59), and the first relation in (4.58). The remaining two relations in (4.58) follow easily, e.g.,

$$\mathrm{E}[X_t Z_0] \;=\; b_t \mathrm{E}[r_0(\zeta_0^2 - 1)Q^2(a + X_0)] + 2\sum_{s_1<0} b_{t-s_1} b_{-s_1} L_{s_1},$$

where

$$\begin{aligned} L_{s_1} &:= \mathrm{E}[r_{s_1}^2 \sum_{s_2<s_1} b_{-s_2} r_{s_2}] \\ &\le \mathrm{E}^{1/2}[r_{s_1}^4]\mathrm{E}^{1/2}\Big[\Big(\sum_{s_2<s_1} b_{-s_2} r_{s_2}\Big)^2\Big] \\ &= O\Big(\Big(\sum_{s_2<s_1} b_{-s_2}^2\Big)^{1/2}\Big) \;=\; O(|s_1|^{d-(1/2)}), \qquad s_1 \to -\infty. \end{aligned}$$

Therefore

$$\mathrm{E}[X_t Z_0] \;=\; O(t^{d-1}) + \sum_{s_1<0} (t - s_1)^{d-1}(-s_1)^{2d-(3/2)} \;=\; o(t^{2d-1}).$$

This proves (4.58).

Next, let us prove (4.54). Recall the decomposition (4.56). Denote $\xi_t := 2aX_t + Z_t$, then (4.56) can be rewritten as $(r_t^2 - \mathrm{E}r_t^2) - \sum_{s<t} b_{t-s}^2 (r_s^2 - \mathrm{E}r_s^2) = \xi_t$, or

$$r_t^2 - \mathrm{E}r_t^2 \;=\; \sum_{i=0}^{\infty} \varphi_i \xi_{t-i}, \qquad t \in \mathbb{Z}, \tag{4.61}$$

where $\varphi_j \geq 0, j \geq 0$ are the coefficients of the power series

$$\Phi(z) := \sum_{j=0}^{\infty} \varphi_j z^j = (1 - \sum_{j=1}^{\infty} b_j^2 z^j)^{-1}, \quad z \in \mathbb{C}, \; |z| < 1$$

given by $\varphi_0 := 1$,

$$\varphi_j \;:=\; b_j^2 + \sum_{0<k<j} \sum_{0<s_1<\cdots<s_k<j} b_{s_1}^2 b_{s_2-s_1}^2 \ldots b_{s_k-s_{k-1}}^2 b_{j-s_k}^2, \quad j \geq 1. \tag{4.62}$$

From (4.51) and Lemma A.1 (see Appendix A) we infer that

$$\varphi_t = O(t^{2d-2}), \qquad t \to \infty, \tag{4.63}$$

in particular, $\Phi(1) = \sum_{t=0}^{\infty} \varphi_t = 1/(1 - B^2) < \infty$ and the r.h.s. of (4.61) is well-defined. Relation (4.58) implies that

$$\gamma_t := \mathrm{Cov}(\xi_0, \xi_t) \sim 4a^2 \mathrm{Cov}(X_0, X_t) \sim 4a^2 \kappa_3^2 t^{2d-1}, \; t \to \infty \tag{4.64}$$

with $\kappa_3^2 = \beta^2 B(d, 1-2d) \mathrm{E}r_0^2$. Let us show that

$$\mathrm{Cov}(r_t^2, r_0^2) \;=\; \sum_{i,j=0}^{\infty} \varphi_i \varphi_j \gamma_{t-i+j} \;\sim\; \Phi^2(1) \gamma_t, \qquad t \to \infty. \tag{4.65}$$

With (4.64) in mind, (4.65) is equivalent to

$$J_t := \sum_{i,j=0}^{\infty} \varphi_i \varphi_j (\gamma_{t-i+j} - \gamma_t) \;=\; o(t^{2d-1}). \tag{4.66}$$

For a large $L > 0$, split $J_t = J'_{t,L} + J''_{t,L}$, where

$$J'_{t,L} := \sum_{i,j>0:|j-i|\leq L} \varphi_i \varphi_j (\gamma_{t-i+j} - \gamma_t), \qquad J''_{t,L} := \sum_{i,j>0:|j-i|>L} \varphi_i \varphi_j (\gamma_{t-i+j} - \gamma_t).$$

Clearly, (4.66) follows from

$$t^{1-2d} J'_{t,L} \;=\; o(1) \quad \forall\; L > 0 \quad \text{and} \quad \lim_{L\to\infty} \limsup_{t\to\infty} t^{1-2d} J''_{t,L} \;=\; 0. \tag{4.67}$$

The first relation in (4.67) is immediate from (4.64) since the latter implies $\gamma_{t+k} - \gamma_t = o(t^{2d-1})$ for any k fixed.

With (4.63) and (4.64) in mind, the second relation in (4.67) follows from

$$\lim_{L\to\infty} \limsup_{t\to\infty} t^{1-2d} \bar{J}_{t,L} \;=\; 0, \tag{4.68}$$

where $\bar{J}_{t,L} := \sum_{i,j>0:|j-i|>L} i^{2d-2} j^{2d-2} (t^{2d-1} + |t+j-i|_+^{2d-1})$ and where $k_+^{2d-1} := \min(1, k^{2d-1}), k \in \mathbb{Z}_+$.

Split the last sum according to whether $|t+j-i| \ge t/2$, or $|t+j-i| < t/2$. Then

$$\begin{aligned} \bar{J}'_{t,L} \;&:=\; \sum_{i,j>0:|j-i|>L, |t+j-i|\ge t/2} i^{2d-2} j^{2d-2} (t^{2d-1} + |t+j-i|^{2d-1}) \\ &\le\; C t^{2d-1} \sum_{i,j>0:|j-i|>L} i^{2d-2} j^{2d-2} \le C t^{2d-1} L^{2d-1} \end{aligned}$$

follows by $\sum_{i,j>0:|j-i|>L} i^{2d-2} j^{2d-2} \le \sum_{0<i<L/2, j>L/2} i^{2d-2} j^{2d-2} + \sum_{i>L/2, j>0} i^{2d-2} j^{2d-2} = O(L^{2d-1})$. Therefore, $\lim_{L\to\infty} \limsup_{t\to\infty} t^{1-2d} \bar{J}'_{t,L} = 0$.

Next, since $|t+j-i| < t/2$ implies $i > t/2$, so with $k := t+j-i$ we obtain

$$\begin{aligned} \bar{J}''_{t,L} \;&\le\; C t^{2d-2} \sum_{i,j>0:|t+j-i|<t/2} j^{2d-2} (t^{2d-1} + |t+j-i|_+^{2d-1}) \\ &\le\; C t^{2d-2} \sum_{j>0} j^{2d-2} \sum_{|k|<t/2} (t^{2d-1} + |k|_+^{2d-1}) \\ &\le\; C t^{4d-2}, \end{aligned}$$

implying $\limsup_{t\to\infty} t^{1-2d} \bar{J}''_{t,L} = 0$ for any $L > 0$. This proves (4.67), (4.66), and (4.65). Clearly, (4.54) follows from (4.65) and (4.64).

It remains to show the invariance principle in (4.55). With (4.61) in mind, decompose $S_n(\tau) := \sum_{t=1}^{[n\tau]} (r_t^2 - \mathrm{E} r_t^2) = \sum_{i=1}^{3} S_{ni}(\tau)$, where

$$\begin{aligned} S_{n1}(\tau) \;&:=\; 2a\Phi(1) \sum_{t=1}^{[n\tau]} X_t, \\ S_{n2}(\tau) \;&:=\; \Phi(1) \sum_{t=1}^{[n\tau]} Z_t, \\ S_{n3}(\tau) \;&:=\; \sum_{t=1}^{[n\tau]} \sum_{i=0}^{\infty} \varphi_i (\xi_{t-i} - \xi_t). \end{aligned}$$

Here, $\mathrm{E}S_{n2}^2(\tau) = o(n^{2d+1})$ follows from (4.58). Consider

$$\mathrm{E}S_{n3}^2(\tau) \;:=\; \sum_{t,s=1}^{[n\tau]} \sum_{i,j=0}^{\infty} \varphi_i \varphi_j \mathrm{E}(\xi_{t-i} - \xi_t)(\xi_{s-j} - \xi_s) \;=\; \sum_{t,s=1}^{[n\tau]} \rho_{t-s},$$

where $\rho_t := \sum_{i,j=0}^{\infty} \varphi_i \varphi_j (\gamma_{t+j-i} - \gamma_{t+j} - \gamma_{t-i} + \gamma_t) = o(t^{2d-1})$ follows similarly to (4.67). Hence, $S_{ni}(\tau) = o_p(n^{-d-1/2})$, $i = 2, 3$. The convergence $n^{-d-1/2} S_{n1}(\tau) \to_{D[0,1]} \kappa_2 B_{d+(1/2)}(\tau)$ follows from Theorem 4.4.1.

This completes the proof of Theorem 4.4.2. $\square$

4.5 Leverage

Given a stationary conditionally heteroscedastic time series $\{r_t\}$ with $\mathrm{E}|r_t|^3 < \infty$ and conditional variance $\sigma_t^2 = \mathrm{Var}(r_t^2 \mid r_s, s < t)$, leverage (a tendency of σ_t^2 to move into the opposite direction as r_s for $s < t$) is usually measured by the covariance $h_{t-s} = \mathrm{Cov}(\sigma_t^2, r_s)$. Following Giraitis et al. (2004), we say that $\{r_t\}$ *has leverage of order* k ($1 \le k < \infty$) (denoted by $\{r_t\} \in \ell(k)$) whenever

$$h_j < 0, \qquad 1 \le j \le k. \tag{4.69}$$

Note that for $\{r_t\}$ in (4.1),

$$h_j = \mathrm{E}[r_j^2 r_0], \qquad j = 0, 1, \dots \tag{4.70}$$

is the mixed moment function. Below, we show that in the case of the quadratic σ_t^2 in (4.4) (correspoding to model (4.53)) and $\mu_3 = \mathrm{E}[\zeta_0^3] = 0$, the function h_j in (4.70) satisfies a linear equation in (4.75), below, which can be analyzed and the leverage effect for $\{r_t\}$ in (4.53) established in spirit of Giraitis et al. (2004).

Let $L^2(\mathbb{Z}_+)$ be the Hilbert space of all real sequences $\psi = (\psi_j, j \in \mathbb{Z}_+), \mathbb{Z}_+ := \{1, 2, \dots\}$ with finite norm $\|\psi\| := (\sum_{j=1}^{\infty} \psi_j^2)^{1/2} < \infty$. As in the previous sections, let $B := \left(\sum_{j=1}^{\infty} b_j^2\right)^{1/2}$ and assume that $\{\zeta_t\}$ is an i.i.d. sequence with zero mean and unit variance; $\mu_i := \mathrm{E}\zeta_0^i$, $i = 1, 2, \dots$.

The following theorem establishes a criterion for the presence or absence of leverage in model (4.53), analogous to the Theorem 2.4 in Giraitis et al. (2004). We also note that the proof of Theorem 4.5.1 is simpler than that of the above mentioned theorem, partly because of the assumption $\mu_3 = 0$ used in the derivation of equation (4.75). Particularly, for the Asymmetric ARCH(1) in (4.29) with $\mathrm{E}|r_0|^3 < \infty, \mu_3 =$

$\mathrm{E}\zeta_0^3 = 0$ the leverage function is $h_j = 2m_2 ab^{2j-1}$, see (4.36), and $\{r_t\} \in \ell(k)$ is equivalent to $ab < 0$. Apparently, conditions $\mu_3 = 0$ and $B^2 < 1/5$ are not necessary for the statement of Theorem 4.5.1 although a similar condition $|\mu_3| \le 2(1-5B^2)/B(1+3B^2)$ appears in the study of the leverage effect in Giraitis et al. (2004), (51).

Theorem 4.5.1 *Let $\{r_t\}$ be a stationary L^2-solution of (4.53) with $\mathrm{E}|r_0|^3 < \infty$, $|\mu|_3 < \infty$. Assume in addition that $B^2 < 1/5$ and $\mu_3 = \mathrm{E}\zeta_0^3 = 0$. Then for any fixed k such that $1 \le k \le \infty$:*

(i) if $ab_1 < 0$, $ab_j \le 0$, $j = 2, \ldots, k$, then $\{r_t\} \in \ell(k)$,

(ii) if $ab_1 > 0$, $ab_j \ge 0$, $j = 2, \ldots, k$, then $h_j > 0$, $j = 1, \ldots, k$.

Proof. Let us first prove that $\|h\| < \infty$. Note that

$$\lim_{n\to\infty} \mathrm{E}\Big(\sum_{-n<s<t} b_{t-s} r_s\Big)^2 r_0 \;=\; \mathrm{E}\Big(\sum_{-\infty<s<t} b_{t-s} r_s\Big)^2 r_0, \tag{4.71}$$

which follows from the definition of L^3-solution of (4.53) and Remark 4.2.2. Then using (4.71), $\mathrm{E}r_t = \mathrm{E}[r_t^3] = \mathrm{E}[r_t r_s] = 0, s < t$ we obtain

$$\begin{aligned} h_j &= \lim_{n\to\infty} \mathrm{E}\Big[\Big(c^2 + a^2 + 2a \sum_{-n<s<t} b_{t-s} r_s + \sum_{-n<s<t} b_{t-s}^2 r_s^2 \\ &\qquad + 2 \sum_{-n<s_2<s_1<t} b_{t-s_1} b_{t-s_2} r_{s_1} r_{s_2}\Big) r_{t-j}\Big] \\ &= 2am_2 b_j + \sum_{t-j<s<t} b_{t-s}^2 h_{j+s-t} + 2b_j \lim_{n\to\infty} \mathrm{E}R_n(t,j), \end{aligned} \tag{4.72}$$

where $R_n(t,j) := r_{t-j}^2 \sum_{-n<s<t-j} b_{t-s} r_s$. Using Hölder's and Rosenthal's (4.13) inequalities we obtain

$$\begin{aligned} |\mathrm{E}R_n(t,j)| &\le \mathrm{E}^{2/3}|r_{t-j}|^3 \mathrm{E}^{1/3}\Big|\sum_{-n<s<t-j} b_{t-s} r_s\Big|^3 \\ &\le \mathrm{E}|r_0|^3 K_3\Big(\sum_{-n<s<t-j} b_{t-s}^2\Big)^{3/2} \;\le\; C. \end{aligned} \tag{4.73}$$

Hence,

$$|h_j| \;\le\; C|b_j| + \sum_{0<i<j} b_{j-i}^2 |h_i| \;\le C\Big(|b_j| + \sum_{0<i<j} \varphi_{j-i}|b_i|\Big), \tag{4.74}$$

where the first inequality in (4.74) follows from (4.72) and (4.73) and the second inequality in (4.74) by iterating the first one with φ_j as in (4.62). Since $\sum_{j=1}^\infty \varphi_j = \sum_{k=1}^\infty B^{2k} = 1/(1-B^2)$, from the second inequality in (4.74) we obtain $\|h\| \le CB/(1-$

$B^2) < \infty$. The last fact implies $\mathrm{E}R_n(t,j) = \sum_{i=1}^{n+t-j} h_i b_{i+j} \to \sum_{i=1}^{\infty} h_i b_{i+j}$. From (4.72) we obtain that the leverage function $h \in L^2(\mathbb{Z}^+)$ is a solution of the linear equation:

$$h_j \;=\; 2ab_j m_2 + \sum_{0<i<j} b_i^2 h_{j-i} + 2b_j \sum_{i>0} b_{i+j} h_i, \qquad j = 1, 2, \ldots. \tag{4.75}$$

From Minkowski's inequality, we get

$$\sum_{j>0} \Big(\sum_{0<i<j} b_i^2 h_{j-i} \Big)^2 \;\le\; B^4 \|h\|^2, \quad \sum_{j>0} \Big(b_j \sum_{i>0} b_{i+j} h_i \Big)^2 \;\le\; B^4 \|h\|^2$$

and then (4.75) implies that $\|h\| \le 2|a|m_2 B + 3B^2 \|h\|$, or

$$\|h\| \;\le\; \frac{2|a|m_2 B}{1 - 3B^2} \tag{4.76}$$

provided $B^2 < 1/3$.

Let us prove the statements (i) and (ii) of Theorem 4.5.1 for $k = 1$. From (4.75) it follows that

$$h_1 \;=\; 2am_2 b_1 + 2b_1 \sum_{u=1}^{\infty} h_u b_{1+u} \;=\; 2b_1 (am_2 + \sum_{u=1}^{\infty} h_u b_{1+u}).$$

Since $|\sum_{u=1}^{\infty} h_u b_{1+u}| \le \|h\| B$, we have $\mathrm{sgn}(h_1) = \mathrm{sgn}(b_1 a)$ provided $\|h\| B < |a| m_2$ holds. The last relation follows from (4.76) and $B^2 < 1/5$; indeed,

$$\|h\| B \le \frac{2|a| m_2 B^2}{1 - 3B^2} \le |a| m_2.$$

This proves (i) and (ii) for $k = 1$.

The general case $k \ge 1$ follows similarly by induction on k. Indeed, from (4.75) we have that

$$h_k \;=\; 2b_k (am_2 + \sum_{u=1}^{\infty} h_u b_{k+u}) + \sum_{j=1}^{k-1} b_{k-j}^2 h_j.$$

Assume $h_1, \ldots, h_{k-1} < 0$, then the second term $\sum_{j=1}^{k-1} b_{k-j}^2 h_j < 0$. Moreover,

$$|\sum_{u=1}^{\infty} h_u b_{k+u}| \le \|h\| B < |a| m_2$$

implying that the sign of the first term is the same as $\mathrm{sgn}(ab_k)$.

Theorem 4.5.1 is proved. $\square$

4.6 A generalized nonlinear model for long memory conditional heteroscedasticity

The present section extends the results of previous sections to a more general class of volatility forms:

$$r_t \;=\; \zeta_t\sigma_t, \qquad \sigma_t^2 \;=\; Q^2\Big(a+\sum_{j=1}^{\infty} b_j r_{t-j}\Big)+\gamma\sigma_{t-1}^2, \tag{4.77}$$

where $\{\zeta_t\}$ are standardized i.i.d. random variables, a, b_j are real parameters, $Q(x)$ is a Lipschitz function of real variable $x \in \mathbb{R}$ and $0 \le \gamma < 1$ is a parameter. For most of the statements below, the proofs are analogous to the proofs of corresponding statements in previous sections and are omitted. The only exception is the proof of Theorem 4.6.2 where a new condition for the existence of stationary solution that does not use the Rosenthal constant is obtained.

A general impression from our results is that the GQARCH modification (corresponding to (4.77) with Q in (4.3)) of the QARCH model discussed in previous sections (see also Doukhan et al. (2016)) allows for a more realistic volatility modeling as compared to the LARCH and QARCH models, at the same time preserving the long memory and the leverage properties of the above mentioned models.

Stationary solution

First we consider the existence of stationary solution of (4.77). Since $0 \le \gamma < 1$, equations (4.77) yield

$$\sigma_t^2 \;=\; \sum_{\ell=0}^{\infty}\gamma^{\ell}Q^2(a+X_{t-\ell}) \quad\text{and}\quad r_t \;=\; \zeta_t\sqrt{\sum_{\ell=0}^{\infty}\gamma^{\ell}Q^2(a+X_{t-\ell})}, \tag{4.78}$$

where

$$X_t \;:=\; \sum_{s<t} b_{t-s} r_s. \tag{4.79}$$

In other words, stationary solution of (4.77), or

$$r_t \;=\; \zeta_t\sqrt{\sum_{\ell=0}^{\infty}\gamma^{\ell}Q^2(a+\sum_{j=1}^{\infty} b_j r_{t-\ell-j})} \tag{4.80}$$

can be defined via (4.79), or stationary solution of

$$X_t \; := \; \sum_{s<t} b_{t-s}\zeta_s \sqrt{\sum_{\ell=0}^{\infty} \gamma^\ell Q^2(a+X_{s-\ell})}, \tag{4.81}$$

and vice versa.

Definition 4.6.1 *Let $p>0$ be an arbitrary real number.*

(i) By L^p-solution of (4.78) *or/and* (4.80) *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* (4.80) *holds.*

(ii) By L^p-solution of (4.81) *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* (4.81) *holds.*

Define

$$B_p := \begin{cases} \sum_{j=1}^{\infty}|b_j|^p, & 0<p<2, \\ \Big(\sum_{j=1}^{\infty} b_j^2\Big)^{p/2}, & p\ge 2, \end{cases} \qquad B_{p,\gamma} := \begin{cases} B_p/(1-\gamma^{p/2}), & 0<p<2, \\ B_p/(1-\gamma)^{p/2}, & p\ge 2. \end{cases} \tag{4.82}$$

Note $B_p = B_{p,0}$.

Proposition 4.6.1 says that equations (4.80) and (4.81) are equivalent in the sense that by solving one the these equations one readily obtains a solution to the other one.

Proposition 4.6.1 *Let Q be a measurable function satisfying* (4.16) *with some $c_i\ge 0$, $i=1,2$ and $\{\zeta_t\}$ 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 $0\le\gamma<1$.*

(i) Let $\{X_t\}$ be a stationary L^p-solution of (4.81) *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* (4.78) *is a stationary L^p-solution of* (4.80) *and*

$$\mathrm{E}|r_t|^p \;\le\; C(1+\mathrm{E}|X_t|^p). \tag{4.83}$$

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

$$\mathrm{E}[r_t|\mathcal{F}_{t-1}]=0, \qquad \mathrm{E}[|r_t|^p|\mathcal{F}_{t-1}]=|\mu|_p\sigma_t^p. \tag{4.84}$$

(ii) Let $\{r_t\}$ be a stationary L^p-solution of (4.80). *Then X_t in* (4.79) *is a stationary*

L^p-solution of (4.81) *such that*

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

Moreover, for $p \ge 2$

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

Remark 4.6.1 Let $p \ge 2$ and $|\mu|_p < \infty$, then by inequality (4.13), $\{r_t\}$ being a stationary L^p-solution of (4.78) is equivalent to $\{r_t\}$ being a stationary L^2-solution of (4.78) with $\mathrm{E}|r_0|^p < \infty$. Similarly, if Q and $\{\zeta_t\}$ satisfy the conditions of Proposition 4.6.1 and $p \ge 2$, then $\{X_t\}$ being a stationary L^p-solution of (4.79) is equivalent to $\{X_t\}$ being a stationary L^2-solution of (4.79) with $\mathrm{E}|X_0|^p < \infty$. See also Section 4.2, Remark 4.2.2.

Theorem 4.6.1 extends Theorem 4.2.1 from $\gamma = 0$ to $\gamma > 0$.

Theorem 4.6.1 *Let $\{\zeta_t\}$ satisfy the conditions of Proposition 4.6.1 and Q satisfy the Lipschitz condition in* (4.15).

(i) Let $p > 0$ and

$$K_p^{1/p} \, |\mu|_p^{1/p} \, \mathrm{Lip}_Q \, B_{p,\gamma}^{1/p} < 1, \tag{4.85}$$

where K_p is the absolute constant from the moment inequality in (4.13). *Then there exists a unique stationary L^p-solution $\{X_t\}$ of* (4.81) *and*

$$\mathrm{E}|X_t|^p \ \le \frac{C(p,Q)|\mu|_p B_p}{1 - K_p|\mu|_p \mathrm{Lip}_Q^p B_{p,\gamma}}, \tag{4.86}$$

where $C(p,Q) < \infty$ depends only on p and c_1, c_2 in (4.16).

(ii) Assume, in addition, that $Q^2(x) = c_1^2 + c_2^2 x^2$, where $c_i \ge 0, i = 1, 2$, 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 $\{X_t\}$ of (4.81) *with $a \ne 0$.*

A major shortcoming of Theorem 4.6.1 (also Theorem 4.2.1) is the presence of the universal constant K_p in the condition (4.85). The upper bound of K_p given in Osękowski (2012) leads to restrictive conditions on $B_{p,\gamma}$ in (4.85) for the existence of L^p-solution, $p > 2$. For example, for $p = 4$ the above mentioned bound in Osękowski (2012) gives

$$K_4 \mu_4 B_2^2/(1-\gamma)^2 \le (27.083)^4 \mu_4 B_2^2/(1-\gamma)^2 < 1 \tag{4.87}$$

requiring $B_2 = \sum_{j=1}^{\infty} b_j^2$ to be very small (see also Remarks 4.2.1, 4.2.3). Since statistical inference based of "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. (2017), also Chapter 5), the question arises to derive less restrictive conditions for the existence of these moments which do not involve the Rosenthal constant K_p. This is achieved in the subsequent Theorem 4.6.2. Particularly, for $\gamma = 0, \mathrm{Lip}_Q = 1$ the sufficient condition (4.89) of Theorem 4.6.2 for the existence of $\mathrm{E}r_t^p, p \ge 2$ even becomes

$$\sum_{j=2}^{p} \binom{p}{j} |\mu_j| \sum_{k=1}^{\infty} |b_k|^j \ < \ 1. \tag{4.88}$$

Condition (4.88) coincides with the corresponding condition in the LARCH case in Giraitis et al. (2004), Proposition 3. Moreover, (4.88) and (4.89) apply to more general classes of ARCH models in (4.77) to which the specific Volterra series techniques used in Giraitis et al. (2000), Giraitis et al. (2004) are not applicable. In the particular case $p = 4$ condition (4.88) 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 (4.87) based on Theorem 4.6.1.

Theorem 4.6.2 *Let $\{\zeta_t\}$ satisfy the conditions of Proposition 4.6.1 and Q satisfy the Lipschitz condition in* (4.15)*.*
Let $p = 2, 4, \ldots$ be even and

$$\sum_{j=2}^{p} \binom{p}{j} |\mu_j| \mathrm{Lip}_Q^j \sum_{k=1}^{\infty} |b_k|^j \ < \ (1-\gamma)^{p/2}. \tag{4.89}$$

Then there exists a unique stationary L^p-solution $\{X_t\}$ of (4.81)*.*

Proof. For $p = 2$, condition (4.89) agrees with $\mathrm{Lip}_Q^2 B_{2,\gamma} < 1$ or condition (4.85) so we shall assume $p \ge 4$ in the subsequent proof. In the latter case (4.89) implies $\mathrm{Lip}_Q^2 B_{2,\gamma} < 1$ and the existence of a stationary L^2-solution $\{X_t\}$ of (4.81). It suffices to show that the above L^2-solution satisfies $\mathrm{E}X_t^p < \infty$.

Towards this end similarly as in the proof of Theorem 4.6.1 (i) consider the solution $\{X_t^{(n)}\}$ of (4.81) with zero initial condition at $t \le -n$ recurently defined as follows

$$X_t^{(n)} \ := \ \begin{cases} 0, & t \le -n, \\ \sum_{s=-n}^{t-1} b_{t-s}\zeta_s\sigma_s^{(n)}, & t > -n, \quad t \in \mathbb{Z}, \end{cases} \tag{4.90}$$

where $\sigma_s^{(n)} := \sqrt{\sum_{\ell=0}^{n+s} \gamma^\ell Q^2(a + X_{s-\ell}^{(n)})}$.

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 (4.89)

$$\mathrm{E}(X_t^{(n)})^p \ < \ C, \tag{4.91}$$

where the constant $C < \infty$ does not depend on t, n.

Since p is even for any $t > -n$ we have that

$$\begin{aligned}
\mathrm{E}(X_t^{(n)})^p \ &= \ \sum_{s_1,\dots,s_p=-n}^{t-1} \mathrm{E}\Big[b_{t-s_1}\zeta_{s_1}\sigma_{s_1}^{(n)} \dots b_{t-s_p}\zeta_{s_p}\sigma_{s_p}^{(n)}\Big] \\
&= \ \sum_{j=2}^{p} \binom{p}{j} \sum_{s=-n}^{t-1} b_{t-s}^j \mu_j \mathrm{E}\Big[(\sigma_s^{(n)})^j \Big(\sum_{u=-n}^{s-1} b_{t-u}\zeta_u \sigma_u^{(n)}\Big)^{p-j}\Big].
\end{aligned} \tag{4.92}$$

Hence using Hölder's inequality:

$$|\mathrm{E}\xi^j\eta^{p-j}| \ \le \ c^j \mathrm{E}^{j/p}|\xi/c|^p \mathrm{E}^{(p-j)/p}|\eta|^p \ \le \ c^j\Big[\frac{j}{pc^p}\mathrm{E}|\xi|^p + \frac{p-j}{p}\mathrm{E}|\eta|^p\Big], \quad 1 \le j \le p, \ c > 0$$

we obtain

$$\begin{aligned}
\mathrm{E}(X_t^{(n)})^p \ &\le \ \sum_{j=2}^{p} \binom{p}{j} |\mu_j| c_3^j \sum_{s=-n}^{t-1} |b_{t-s}^j| \Big\{ \tfrac{j}{pc_3^p} \mathrm{E}(\sigma_s^{(n)})^p + \tfrac{p-j}{p} \mathrm{E}\Big(\sum_{u=-n}^{s-1} b_{t-u}\zeta_u\sigma_u^{(n)}\Big)^p \Big\} \\
&= \ \sum_{s=-n}^{t-1} \beta_{1,t-s} \mathrm{E}(\sigma_s^{(n)}/c_3)^p + \sum_{s=-n}^{t-1} \beta_{2,t-s} \mathrm{E}\Big(X_{t,s}^{(n)}\Big)^p,
\end{aligned} \tag{4.93}$$

where $X_{t,s}^{(n)} := \sum_{u=-n}^{s-1} b_{t-u}\zeta_u\sigma_u^{(n)}$, $c_3 > \mathrm{Lip}_Q$ and where

$$\beta_{1,t-s} \ := \ \sum_{j=2}^{p} \frac{j}{p} \binom{p}{j} |b_{t-s}^j||\mu_j| c_3^j, \qquad \beta_{2,t-s} \ := \ \sum_{j=2}^{p} \frac{p-j}{p} \binom{p}{j} |b_{t-s}^j||\mu_j| c_3^j.$$

The last expectation in (4.93) can be evaluated similarly to (4.92)-(4.93):

$$\begin{aligned}
\mathrm{E}\Big(X_{t,s}^{(n)}\Big)^p \ &= \ \sum_{j=2}^{p} \binom{p}{j} \sum_{u=-n}^{s-1} b_{t-u}^j \mu_j \mathrm{E}\Big[(\sigma_u^{(n)})^j \Big(\sum_{v=-n}^{u-1} b_{t-v}\zeta_v\sigma_v^{(n)}\Big)^{p-j}\Big] \\
&\le \ \sum_{u=-n}^{s-1} \beta_{1,t-u} \mathrm{E}(\sigma_u^{(n)}/c_3)^p + \sum_{u=-n}^{s-1} \beta_{2,t-u} \mathrm{E}\Big(X_{t,u}^{(n)}\Big)^p.
\end{aligned}$$

Proceeding recurrently with the above evaluation results in the inequality:

$$\mathrm{E}(X_t^{(n)})^p \;\le\; \sum_{s=-n}^{t-1} \widetilde{\beta}_{t-s}\mathrm{E}(\sigma_s^{(n)}/c_3)^p, \tag{4.94}$$

where

$$\widetilde{\beta}_{t-s} \;:=\; \beta_{1,t-s}\Big(1+\sum_{k=1}^{t-s-1}\sum_{s<u_k<\cdots<u_1<t}\beta_{2,t-u_1}\ldots\beta_{2,t-u_k}\Big).$$

Let $\beta_i := \sum_{t=1}^\infty \beta_{i,t}, i = 1, 2,\ \widetilde{\beta} := \sum_{t=1}^\infty \widetilde{\beta}_t$. By assumption (4.89),

$$\beta_1+\beta_2 \;=\; \sum_{j=2}^{p}\binom{p}{j}|\mu_j|c_3^j\sum_{k=1}^\infty |b_k|^j \;<\; (1-\gamma)^{p/2}$$

whenever $(c_3 - \mathrm{Lip}_Q) > 0$ is small enough, and therefore

$$\begin{aligned}\frac{\widetilde{\beta}}{(1-\gamma)^{p/2}} \;&\le\; \frac{1}{(1-\gamma)^{p/2}}\sum_{t=1}^\infty \beta_{1,t}\Big(1+\sum_{k=1}^\infty \beta_2^k\Big)\\ &=\; \frac{1}{(1-\gamma)^{p/2}}\frac{\beta_1}{1-\beta_2} \;<\; 1.\end{aligned} \tag{4.95}$$

Next, let us estimate the expectation on the r.h.s. of (4.94) in terms of the expectations on the l.h.s. Using (4.16) and Minkowski's inequalities we obtain

$$\begin{aligned}\mathrm{E}^{2/p}(\sigma_s^{(n)})^p \;&\le\; \sum_{\ell=0}^{s+n}\gamma^\ell \mathrm{E}^{2/p}|Q(a+X_{s-\ell}^{(n)}|^p\\ &\le\; \sum_{\ell=0}^{s+n}\gamma^\ell \mathrm{E}^{2/p}|c_1^2+c_2^2(a+X_{s-\ell}^{(n)})^2|^{p/2}\\ &\le\; C + c_3^2\sum_{\ell=0}^{n+s}\gamma^\ell \mathrm{E}^{2/p}(X_{s-\ell}^{(n)})^p,\end{aligned}$$

where $c_3 > c_2 > \mathrm{Lip}_Q$ and $c_3-\mathrm{Lip}_Q > 0$ can be chosen arbitrarily small. Particularly, for any fixed $T \in \mathbb{Z}$

$$\sup_{-n\le s<T}\mathrm{E}^{2/p}(\sigma_s^{(n)})^p \;\le\; \frac{c_3^2}{(1-\gamma)}\sup_{-n\le s<T}\mathrm{E}^{2/p}(X_s^{(n)})^p + C.$$

Substituting the last bound into (4.94) we obtain

$$\sup_{-n\le t<T}\mathrm{E}^{2/p}(X_t^{(n)})^p \;\le\; \frac{\widetilde{\beta}^{2/p}}{(1-\gamma)}\sup_{-n\le s<T}\mathrm{E}^{2/p}(X_s^{(n)})^p + C. \tag{4.96}$$

Relations (4.96) and (4.95) imply

$$\sup_{-n\le t<T} \mathrm{E}^{2/p}(X_t^{(n)})^p \le \frac{C}{1-\frac{\widetilde{\beta}^{2/p}}{(1-\gamma)}} \;<\; \infty$$

proving (4.91) and the theorem, too. $\square$

Example 4.6.1 (Asymmetric GARCH(1,1)) The asymmetric GARCH(1,1) model of Engle (1990) corresponds to

$$\sigma_t^2 = c^2 + (a + br_{t-1})^2 + \gamma\sigma_{t-1}^2, \tag{4.97}$$

or

$$\sigma_t^2 = \theta + \psi r_{t-1} + a_{11} r_{t-1}^2 + \delta\sigma_{t-1}^2 \tag{4.98}$$

in the parametrization of Sentana (1995), (5), with parameters in (4.97), (4.98) related by

$$\theta = c^2 + a^2, \quad \delta = \gamma, \quad \psi = 2ab, \quad a_{11} = b^2. \tag{4.99}$$

Under the conditions 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 (1995)) process $\{r_t\}$ with finite variance and $a \ne 0$ exists if and only if $B_{2,\gamma} = b^2/(1-\gamma) < 1$, or

$$b^2 + \gamma < 1, \tag{4.100}$$

see Theorem 4.6.1 (ii). Condition (4.100) agrees with condition $a_{11} + \delta < 1$ for covariance stationarity in Sentana (1995). Under the assumptions that the distribution of ζ_t is symmetric and $\mu_4 = \mathrm{E}\zeta_t^4 < \infty$, Sentana (1995) provides a sufficient condition for finiteness of $\mathrm{E}r_t^4$ together with explicit formula

$$\mathrm{E}r_t^4 \;=\; \frac{\mu_4\theta[\theta(1 + a_{11} + \delta) + \psi^2]}{(1 - a_{11}^2\mu_4 - 2a_{11}\delta - \delta^2)(1 - a_{11} - \delta)}. \tag{4.101}$$

The sufficient condition of Sentana (1995) for $\mathrm{E}r_t^4 < \infty$ is $\mu_4 a_{11}^2 + 2a_{11}\delta + \delta^2 < 1$, which translates to

$$\mu_4 b^4 + 2b^2\gamma + \gamma^2 < 1 \tag{4.102}$$

in terms of the parameters of (4.97). Condition (4.102) seems weaker than the sufficient condition $\mu_4 b^4 + 6b^2 < (1-\gamma)^2$ of Theorem 4.6.2 for the existence of L^4-solution of (4.97).

Following the approach in Section 4.2, Example 4.2.2, below we find explicitly the covariance function $\rho(t) := \mathrm{Cov}(r_0^2, r_t^2)$, including the expression in (4.101), for stationary solution of the asymmetric GARCH(1,1) in (4.97). We can write the following moment equations:

$$
\begin{aligned}
m_2 &= (c^2+a^2)/(1-b^2-\gamma), \quad m_3(0)=0, \\
m_3(1) &= \sum_{\ell=0}^{\infty} \gamma^\ell \mathrm{E}(c^2+a^2+2abr_{-\ell}+b^2 r_{-\ell}^2) r_0 = 2abm_2, \\
m_3(t) &= \sum_{\ell=0}^{\infty} \gamma^\ell \mathrm{E}(c^2+a^2+2abr_{t-\ell-1}+b^2 r_{t-\ell-1}^2) r_0 \\
&= 2abm_2\gamma^{t-1} + b^2 \sum_{\ell=0}^{t-2} \gamma^\ell m_3(t-\ell-1), \quad t \ge 2. \qquad (4.103)
\end{aligned}
$$

From equations above one can show by induction that $m_3(t) = 2abm_2(\gamma+b^2)^{t-1}, t \ge 1$.

Similarly,

$$
\begin{aligned}
m_4(0) &= \mu_4 \mathrm{E}((c^2+a^2)+2abr_0+b^2r_0^2+\gamma\sigma_0^2)^2 \\
&= \mu_4\Big((c^2+a^2)^2+(2ab)^2m_2+b^4m_4(0)+2(c^2+a^2)(b^2+\gamma)m_2 \\
&+ (2b^2\gamma+\gamma^2)m_4(0)/\mu_4\Big), \\
m_4(t) &= \sum_{\ell=0}^{\infty} \gamma^\ell \mathrm{E}(c^2+a^2+2abr_{t-\ell-1}+b^2r_{t-\ell-1}^2)r_0^2 \\
&= \sum_{\ell=0}^{\infty} \gamma^\ell (c^2+a^2)m_2 + b^2\sum_{\ell=0}^{\infty}\gamma^\ell m_4(|t-\ell-1|) \\
&+ 2ab\sum_{\ell=t}^{\infty}\gamma^\ell m_3(\ell-t+1), \quad t \ge 1.
\end{aligned}
$$

Using $2ab\sum_{\ell=t}^{\infty}\gamma^\ell m_3(\ell-t+1) = 4a^2b^2m_2\sum_{\ell=t}^{\infty}\gamma^\ell(\gamma+b^2)^{\ell-t} = 4a^2b^2m_2\gamma^t/(1-\gamma(\gamma+b^2))$ and $\rho(t) = m_4(t)-m_2^2$ we obtain the system of equations

$$
\begin{aligned}
\rho(0) &= m_4(0)-m_2^2, \\
\rho(t) &= b^2\sum_{\ell=0}^{\infty}\gamma^\ell\rho(|t-\ell-1|) + 4a^2b^2m_2\gamma^t/(1-\gamma(\gamma+b^2)) \\
&= b^2\sum_{\ell=0}^{t-2}\gamma^\ell\rho(t-\ell-1) + C\gamma^{t-1}, \quad t \ge 1, \qquad (4.104)
\end{aligned}
$$

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

$$m_4(0) \;=\; \frac{\mu_4 m_2}{1-b^4\mu_4-(2b^2\gamma+\gamma^2)}\Big((c^2+a^2)(1+b^2+\gamma)+(2ab)^2\Big). \quad (4.105)$$

Note that the expression above coincides with (4.101) given that the relations in (4.99) hold.

Since the equation in (4.104) is analogous to (4.103), the solution to (4.104) 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 (4.105) and $\rho(t)=C(\gamma+b^2)^{t-1}, t\geq 1$ giving explicitly the covariances of process $\{r_t^2\}$.

Model properties

The present section studies long memory and leverage properties of the generalized quadratic ARCH (GQARCH) model in (4.77) corresponding to Q in (4.3), viz.,

$$r_t \;=\; \zeta_t\sqrt{\sum_{\ell=0}^{\infty}\gamma^\ell\Big(c^2+\Big(a+\sum_{s<t-\ell}b_{t-\ell-s}r_s\Big)^2\Big)}, \qquad t\in\mathbb{Z}, \quad (4.106)$$

where $0\leq\gamma<1, a\neq 0, c$ are real parameters, $\{\zeta_t\}$ are standardized i.i.d. random variables, with zero mean and unit variance, and $b_j, j\geq 1$ are real numbers.

Theorem 4.6.3 extends the results on long memory in Theorem 4.4.2 corresponding to $\gamma=0$ to the case $\gamma>0$. In Theorem 4.6.3 and below, $0\leq\gamma<1$, $B_2=\sum_{j=1}^\infty b_j^2$ and $B(\cdot,\cdot)$ is beta function.

Theorem 4.6.3 *Let $\{r_t\}$ be a stationary L^2-solution of* (4.106) *with coefficients b_j decaying regularly as in* (4.51). *Assume in addition that $\mu_4=\mathrm{E}[\zeta_0^4]<\infty$, and $\mathrm{E}[r_t^4]<\infty$. Then*

$$\mathrm{Cov}(r_0^2,r_t^2) \;\sim\; \kappa_1^2 t^{2d-1}, \qquad t\to\infty, \quad (4.107)$$

where $\kappa_1^2:=\left(\frac{2a\beta}{1-\gamma-B_2}\right)^2 B(d,1-2d)\mathrm{E}r_0^2$. Moreover,

$$n^{-d-1/2}\sum_{t=1}^{[n\tau]}(r_t^2-\mathrm{E}r_t^2) \;\to_{D[0,1]}\; \kappa_2 W_{d+(1/2)}(\tau), \qquad n\to\infty, \quad (4.108)$$

where $W_{d+(1/2)}$ is a fractional Brownian motion with Hurst parameter $H = d+(1/2) \in (1/2, 1)$ and $\kappa_2^2 := \kappa_1^2/(d(1+2d))$.

Proposition 4.6.2 below extends to the GQARCH model the leverage effect discussed in Theorem 4.5.1 and Giraitis et al. (2004). The study of leverage for model (4.53) (corresponding to (4.106) with $\gamma = 0$) was based on linear equation for leverage function in (4.75). A similar equation (4.109) for leverage function can be derived for model (4.106) in the general case $0 \le \gamma < 1$. Namely, using $\mathrm{E}r_s = 0$, $\mathrm{E}r_s r_0 = m_2 \mathbf{1}(s=0), \mathrm{E}r_s^2 r_0 = 0\,(s \le 0), \mathrm{E}r_0 r_{s_1} r_{s_2} = \mathbf{1}(s_1 = 0) h_{-s_2}$ $(s_2 < s_1)$ as in proof of Theorem (4.5.1) we have that

$$\begin{aligned}
h_t &= \mathrm{E}r_t^2 r_0 = \sum_{\ell=0}^{t-1} \gamma^\ell \mathrm{E}\Big[(c^2 + (a + \sum_{s<t-\ell} b_{t-\ell-s} r_s)^2) r_0\Big] \\
&= \sum_{\ell=0}^{t-1} \gamma^\ell \Big(2am_2 b_{t-\ell} + \sum_{s<t-\ell} b_{t-\ell-s}^2 \mathrm{E}[r_s^2 r_0]\Big) \\
&+ 2\sum_{\ell=0}^{t-1} \gamma^\ell \sum_{s_2<s_1<t-\ell} b_{t-\ell-s_1} b_{t-\ell-s_2} \mathrm{E}[r_{s_1} r_{s_2} r_0] \\
&= 2am_2 b_{t,\gamma} + \sum_{0<i<t} h_i \tilde{b}_{t-i,\gamma}^2 + 2\sum_{i>0} h_i w_{i,t,\gamma},
\end{aligned} \tag{4.109}$$

where $b_{t,\gamma}, \tilde{b}_{t,\gamma}^2$ are defined as follows

$$b_{t,\gamma} := \sum_{j=0}^{t-1} \gamma^j b_{t-j}, \qquad \tilde{b}_{t,\gamma}^2 := \sum_{j=0}^{t-1} \gamma^j b_{t-j}^2, \qquad t \ge 1 \tag{4.110}$$

and $w_{i,t,\gamma} := \sum_{\ell=0}^{t-1} \gamma^\ell b_{t-\ell} b_{i+t-\ell}$. From now on the proof is analogous to the proof of Theorem 4.5.1 and is not included.

Proposition 4.6.2 *Let $\{r_t\}$ be a stationary L^2-solution of (4.106) 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 $1 \le k \le \infty$:*

(i) if $ab_1 < 0$, $ab_j \le 0, j = 2, \ldots, k$, then $\{r_t\} \in \ell(k)$,

(ii) if $ab_1 > 0$, $ab_j \ge 0, j = 2, \ldots, k$, then $h_j > 0$, for $j = 1, \ldots, k$.

4.7 A simulation study

The (asymmetric) GQARCH model of (4.106) and the LARCH model of (2.11) have similar long memory and leverage properties and both can be used for modelling of financial data with the above properties. The main disadvantage of the latter model vs. the former one seems to be the fact the volatility σ_t may take negative values and is not separated from below by positive constant $c > 0$ as in the case of (4.106).

Consistent QML estimation for 5-parametric long memory GQARCH model in (4.106) with $c > 0$ and $b_j = \beta j^{d-1}$ is discussed in Chapter 5. The parametric form $b_j = \beta j^{d-1}$ of the moving-average coefficients in (4.106) is the same as in Beran and Schützner (2009) for the LARCH model. It is of interest to compare QML estimates and volatility graphs of the GQARCH and LARCH models based on real data. The comparisons are extended to the classical GARCH(1,1) model

$$r_t = \sigma_t \zeta_t, \quad \sigma_t = \sqrt{\omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2}. \tag{4.111}$$

We fit four data generating processes (DGP):

$$\begin{aligned}
(L):& \quad \text{LARCH of (5.2)}, \\
(Q1):& \quad \text{QARCH of (4.106) with } \gamma = 0, \\
(Q2):& \quad \text{QARCH of (4.106) with } \gamma > 0, \\
(G):& \quad \text{GARCH(1,1) of (4.111)},
\end{aligned} \tag{4.112}$$

with $b_j = \beta j^{d-1}$ to daily returns of GSPC (SP500) from 2010 01 01 till 2015 01 01 with $n = 1257$ observations in total. The first three models (L), (Q1), (Q2) have long memory and (G) is short memory. The parameters

$$\begin{aligned}
(L):& \quad (a, \beta, d) = (0.0101, -0.1749, 0.3520), \\
(Q1):& \quad (a, c, \beta, d) = (0.0058, -0.0101, 0.2099, 0.4648), \\
(Q2):& \quad (a, c, \beta, d, \gamma) = (0.0020, -0.0049, 0.2394, 0.2393, 0.7735), \\
(G):& \quad (\omega, \alpha, \beta) = (0.00001, 0.1306, 0.8346),
\end{aligned}$$

are obtained by minimizing the corresponding approximate log-likelihood functions and using the constrains for coefficients to ensure the second order stationarity. The details of the estimation are presented in Chapter 5.

Figure 4.1 presents estimated trajectories of σ_t of four DGP in (4.112), corresponding to the returns of GSPC (original returns plotted at the bottom graph). Observe that the variability of volatility decreases from top to bottom, (Q2) resembling (G) (GARCH(1,1)) trajectory more closely than (L) and (Q1). The graph (Q1) exhibits very sharp peaks and clustering and a tendency to concentrate near the lower threshold c outside of high volatility regions. This unrealistic "threshold effect" is

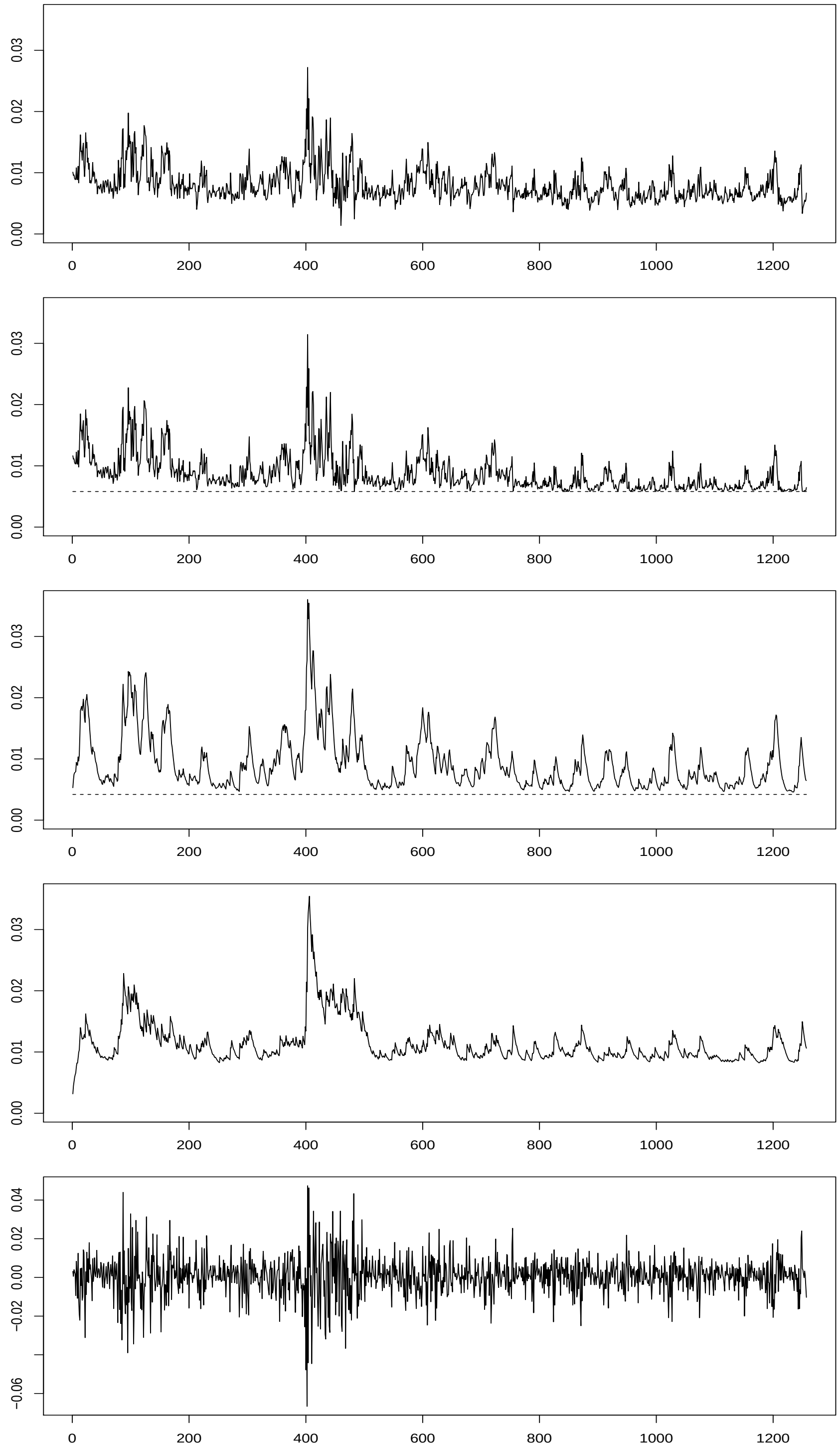

Figure 4.1: Trajectory of DGP: From top to bottom: (L), (Q1), (Q2), (G) and returns of GSPC. The dashed line in (Q1) and (Q2) indicates the threshold $c/\sqrt{1-\gamma} > 0$ in (4.106).

much less pronounced in (Q2) (and also in the other two DGP), due to presence of the autoregressive parameter $\gamma > 0$ which also prevents sharp changes and excessive variability of volatility series. Figure 4.2 illustrates the effect of γ on the marginal distribution of (Q2): with γ increasing, the distribution becomes less skewed and spreads to the right, indicating a lower degree of volatility clustering.

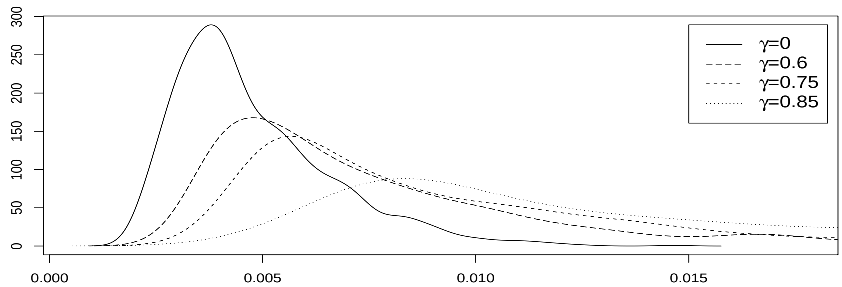

Figure 4.2: Smoothed histograms of DGP (Q2) for different values of γ.

Chapter 5

Quasi-MLE for quadratic ARCH model with long memory

Abstract. We discuss parametric quasi-maximum likelihood estimation for quadratic ARCH (QARCH) process with long memory introduced in Chapter 4 with conditional variance involving the square of inhomogeneous linear combination of observable sequence with square summable weights. The aforementioned model extends the QARCH model of Sentana (1995) and the Linear ARCH model of Robinson (1991) to the case of strictly positive conditional variance. We prove consistency and asymptotic normality of the corresponding QML estimators, including the estimator of long memory parameter $0 < d < 1/2$. A simulation study of empirical MSE is included.

5.1 Introduction

Chapter 4 discussed a class of quadratic ARCH models of the form

$$r_t \; = \; \zeta_t \sigma_t, \qquad \sigma_t^2 \; = \; \omega^2 + \Big(a + \sum_{j=1}^{\infty} b_j r_{t-j}\Big)^2 + \gamma \sigma_{t-1}^2, \tag{5.1}$$

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 $\gamma, \omega, a, b_j, j \geq 1$ are real parameters satisfying certain conditions presented in Theorem 4.6.2. In Chapter 4, (5.1) was called the generalized quadratic ARCH (GQARCH) model. By iterating the second equation in (5.1), the squared volatility in (5.1) 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 $r_{t-1}, r_{t-2}, \ldots$, and hence it represents a particular case of Quadratic ARCH model by Sentana (1995) with $p = \infty$. The model (5.1) includes the classical Asymmetric GARCH(1,1) process of Engle (1990) and the Linear ARCH (LARCH) model of Robinson (1991):

$$r_t \;=\; \zeta_t \sigma_t, \qquad \sigma_t \;=\; a + \sum_{j=1}^{\infty} b_j r_{t-j}. \tag{5.2}$$

The main interest in (5.1) and (5.2) seems 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 in order that the series $\sum_{j=1}^{\infty} b_j r_{t-j}$ converges. Giraitis et al. (2000) proved that the squared stationary solution $\{r_t^2\}$ of the LARCH model in (5.2) with b_j decaying as $j^{d-1}, 0 < d < 1/2$ may have long memory autocorrelations. For the GQARCH model in (5.1), similar results were established in Chapter 4. Namely, assume that the parameters $\gamma, \omega, a, b_j, j \geq 1$ in (5.1) satisfy

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

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

$$6B_2 + 4|\mu_3| \sum_{j=1}^{\infty} |b_j|^3 + \mu_4 \sum_{j=1}^{\infty} b_j^4 \;<\; (1-\gamma)^2, \tag{5.3}$$

where $\mu_p := \mathrm{E}\zeta_0^p,\ p = 1, 2, \ldots,\ B_2 := \sum_{j=1}^{\infty} b_j^2$. Then (see Chapter 4, Theorems 4.6.2 and 4.6.3) there exists a stationary solution of (5.1) 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

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

where $W_{d+(1/2)}$ is a fractional Brownian motion with Hurst parameter $H = d + (1/2) \in (1/2, 1)$ and $\kappa_i > 0, i = 1, 2$ are some constants.

As noted in Chapter 4, the GQARCH model of (5.1) and the LARCH model of (5.2) have similar long memory and leverage properties and both can be used for modelling of financial data with the above properties. The main disadvantage of the latter model vs. the former one seems to be the fact that volatility σ_t in (5.2) may assume negative values and is not separated from below by positive constant $c > 0$ as in the case of (5.1). The standard quasi-maximum likelihood (QML) approach to estimation of LARCH parameters is inconsistent and other estimation methods were developed in Beran and Schützner (2009), Francq and Zakoian (2010b), Levine et al. (2009), Truquet (2014).

The present chapter discusses QML estimation for the 5-parametric GQARCH model

$$\sigma_t^2(\theta) \;=\; \sum_{\ell=0}^{\infty} \gamma^\ell \Big\{\omega^2 + \Big(a + c\sum_{j=1}^{\infty} j^{d-1} r_{t-\ell-j}\Big)^2\Big\}, \tag{5.4}$$

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 (5.4) is the same as in Beran and Schützner (2009) for the LARCH model. Similarly as in Beran and Schützner (2009) we discuss the QML estimator $\widehat{\theta}_n := \arg\min_{\theta\in\Theta} L_n(\theta)$, $L_n(\theta) := \frac{1}{n}\sum_{t=1}^n \left(\frac{r_t^2}{\sigma_t^2(\theta)} + \log \sigma_t^2(\theta)\right)$ involving exact conditional variance in (5.4) 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 (5.4) by $\widetilde{\sigma}_t^2(\theta)$ depending only $r_s, 1 \le s < t$ (see Section 5.2 for the definition). It should be noted that the QML function in Beran and Schützner (2009) is modified to avoid the degeneracy of σ_t^{-1} in (5.2), by introducing an additional tuning parameter $\epsilon > 0$ which affects the performance of the estimator and whose choice is a non-trivial task. For the GQARCH model (5.4) with $\omega > 0$ the above degeneracy problem does not occur and we deal with unmodified QMLE in contrast to Beran and Schützner (2009). We also note that our proofs use different techniques from Beran and Schützner (2009). Particularly, the method of orthogonal Volterra expansions of the LARCH model used in Beran and Schützner (2009) is not applicable for model (5.4), see Chapter 4, Example 4.2.1.

This chapter is organized as follows. In Section 5.2 we define several QML estimators of parameter θ in (5.4). Section 5.3 presents the main results of the paper devoted to consistency and asymptotic normality of the QML estimators. The proofs in this section are based on four lemmas presented in Section 5.4. Finite sample performance of these estimators is investigated in the simulation study in Section 5.5.

5.2 QML estimators

Let $\mathcal{F}_t = \sigma(\zeta_s, s \le t), t \in \mathbb{Z}$ be the sigma-field generated by $\zeta_s, s \le t$. For real $p \ge 2$, define as in Chapter 4 Section 4.6

$$B_p := \Big(\sum_{j=1}^{\infty} b_j^2\Big)^{p/2}, \qquad B_{p,\gamma} := B_p/(1-\gamma)^{p/2}. \tag{5.5}$$

The following assumptions on the parametric GQARCH model in (5.4) are imposed.

Assumption (A): $\{\zeta_t\}$ 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 (5.1) 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 (4.89) in Theorem 4.6.2 with $p = 2$. The QML estimator of $\theta \in \Theta$ is defined as

$$\widehat{\theta}_n := \arg\min_{\theta \in \Theta} L_n(\theta), \tag{5.6}$$

where

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{r_t^2}{\sigma_t^2(\theta)} + \log \sigma_t^2(\theta) \right) \tag{5.7}$$

and $\sigma_t^2(\theta)$ is defined in (5.4), viz.,

$$\begin{aligned} \sigma_t^2(\theta) &= \sum_{\ell=0}^{\infty} \gamma^{\ell} \Big\{ \omega^2 + \Big(a + c Y_{t-\ell}(d) \Big)^2 \Big\}, \qquad \text{where} \\ Y_t(d) &:= \sum_{j=1}^{\infty} j^{d-1} r_{t-j}. \end{aligned} \tag{5.8}$$

Note the definitions in (5.6)-(5.8) depend on (unobserved) $r_s, s \le 0$ and therefore the estimator in (5.6) is usually referred to as the QMLE given infinite past (Beran and Schützner (2009)). A more realistic version of (5.6) is defined as

$$\widetilde{\theta}_n := \arg\min_{\theta \in \Theta} \widetilde{L}_n(\theta), \tag{5.9}$$

where

$$\widetilde{L}_n(\theta) \;:=\; \frac{1}{n}\sum_{t=1}^{n}\Big(\frac{r_t^2}{\widetilde{\sigma}_t^2(\theta)} + \log\widetilde{\sigma}_t^2(\theta)\Big), \quad \text{where} \tag{5.10}$$

$$\widetilde{\sigma}_t^2(\theta) \;:=\; \sum_{\ell=0}^{t-1}\gamma^\ell\Big\{\omega^2 + \Big(a + c\widetilde{Y}_{t-\ell}(d)\Big)^2\Big\}, \quad \widetilde{Y}_t(d) \;:=\; \sum_{j=1}^{t-1} j^{d-1} r_{t-j}.$$

Note all quantities in (5.10) depend only on $r_s, 1 \le t \le n$, hence (5.9) is called the QMLE given finite past. The QML functions in (5.7) and (5.10) 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

$$l_t(\theta) := \frac{r_t^2}{\sigma_t^2(\theta)} + \log\sigma_t^2(\theta), \qquad \widetilde{l}_t(\theta) := \frac{r_t^2}{\widetilde{\sigma}_t^2(\theta)} + \log\widetilde{\sigma}_t^2(\theta). \tag{5.11}$$

Finally, following Beran and Schützner (2009) we define a truncated version of (5.9) involving the last $O(n^\beta)$ quasi-likelihoods $\widetilde{l}_t(\theta), n - [n^\beta] < t \le n$, as follows:

$$\widetilde{\theta}_n^{(\beta)} := \arg\min_{\theta\in\Theta}\widetilde{L}_n^{(\beta)}(\theta), \qquad \widetilde{L}_n^{(\beta)}(\theta) \;:=\; \frac{1}{[n^\beta]}\sum_{t=n-[n^\beta]+1}^{n}\widetilde{l}_t(\theta). \tag{5.12}$$

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 (5.8) and (5.10) 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 $\sigma_t^2(\theta), \widetilde{\sigma}_t^2(\theta), l_t(\theta), \widetilde{l}_t(\theta), L_n(\theta), \widetilde{L}_n(\theta), \widetilde{L}_n^{(\beta)}(\theta)$ etc. are all infinitely differentiable w.r.t. $\theta \in \Theta_0$ a.s. We use the notation

$$L(\theta) := \mathrm{E}L_n(\theta) = \mathrm{E}l_t(\theta) \tag{5.13}$$

and

$$A(\theta) \;:=\; \mathrm{E}\left[\nabla^T l_t(\theta)\nabla l_t(\theta)\right] \qquad \text{and} \qquad B(\theta) \;:=\; \mathrm{E}\left[\nabla^T\nabla l_t(\theta)\right], \tag{5.14}$$

where $\nabla = (\partial/\partial\theta_1, \ldots, \partial/\partial\theta_5)$ and the superscript T stands for transposed vector. Particularly, $A(\theta)$ and $B(\theta)$ are 5×5-matrices. By Lemma 5.4.1, the expectations in (5.14) are well-defined for any $\theta \in \Theta$ under condition $\mathrm{E}r_0^4 < \infty$. We have

$$B(\theta) \;=\; \mathrm{E}[\sigma_t^{-4}(\theta)\nabla^T\sigma_t^2(\theta)\nabla\sigma_t^2(\theta)] \quad \text{and} \quad A(\theta) = \kappa_4 B(\theta), \tag{5.15}$$

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

5.3 Main results

Everywhere in this section $\{r_t\}$ is a stationary solution of model (5.4) as defined in Definition 4.6.1 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 $\Theta \subset \mathbb{R}^5$.

The asymptotic results in Theorems 5.3.1 and 5.3.2 are similar to the results of Beran and Schützner (2009), Theorems 1-4, pertaining to the 3-parametric LARCH model in (5.2) with $b_j = cj^{d-1}$, except that Beran and Schützner (2009) deal with a modified QMLE involving a "tuning parameter" $\epsilon > 0$. As explained in Beran and Schützner (2009), Section 3.2, the convergence rate of $\nabla \widetilde{L}_n(\theta_0)$ and $\widetilde{\theta}_n$ (based on nonstationary truncated observable series in (5.10)) is apparently too slow to guarantee asymptotic normality, this fact being a consequence of long memory in volatility and the main reason for introducing the estimators $\widetilde{\theta}_n^{(\beta)}$ in (5.12). Theorems 5.3.1 and 5.3.2 are based on subsequent Lemmas 5.4.1-5.4.4 which describe properties of the likelihood processes defined in (5.7), (5.10) and (5.11). As noted in Section 5.1, our proofs use different techniques from Beran and Schützner (2009) which rely on explicit Volterra series representation of stationary solution of the LARCH model.

Theorem 5.3.1 *(i) Let* $\mathrm{E}|r_t|^3 < \infty$. *Then* $\widehat{\theta}_n$ *in* (5.6) *is a strongly consistent estimator of* θ_0, *i.e.*

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

(ii) Let $\mathrm{E}|r_t|^5 < \infty$. *Then* $\widehat{\theta}_n$ *in* (5.6) *is asymptotically normal:*

$$n^{1/2}\Big(\widehat{\theta}_n - \theta_0\Big) \overset{law}{\to} N(0, \Sigma(\theta_0)), \tag{5.16}$$

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* (5.15).

Proof. (i) Follows from Lemmas 5.4.2 and 5.4.4 (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 $\theta_n^* \to_{\mathrm{p}} \theta_0$ since $\widehat{\theta}_n \to_{\mathrm{p}} \theta_0$. Then $\nabla^T \nabla L_n(\theta_n^*) \to_{\mathrm{p}} \nabla^T \nabla L(\theta_0)$ by Lemma 5.4.4 (5.57). 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 $n^{1/2}\nabla L_n(\theta_0) \overset{law}{\to} N(0, A(\theta_0))$ follows by the martingale central limit theorem in (Billingsley (1968), Therem 23.1). Then (5.16) follows by Slutsky's theorem and (5.14). $\square$

The following theorem gives asymptotic properties of "finite past" estimators $\widetilde{\theta}_n$ and $\widetilde{\theta}_n^{(\beta)}$ defined in (5.9) and (5.12), respectively.

Theorem 5.3.2 *(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*

$$n^{\beta/2}(\widetilde{\theta}_n^{(\beta)} - \theta_0) \;\overset{law}{\to}\; N(0, \Sigma(\theta_0)), \tag{5.17}$$

where $\Sigma(\theta_0)$ *is the same as in Theorem 5.3.1.*

Proof. Part (i) follows from Lemmas 5.4.2 and 5.4.4 (i) as in the case of Theorem 5.3.1 (i).

To prove part (ii), by Taylor's expansion

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

where $\widetilde{\theta}_n^* \to_{\mathrm{p}} \theta_0$ since $\widetilde{\theta}_n^{(\beta)} \to_{\mathrm{p}} \theta_0$. Then $\nabla^T\nabla\widetilde{L}_n^{(\beta)}(\theta_n^*) \to_{\mathrm{p}} \nabla^T\nabla L(\theta_0)$ by Lemma 5.4.4 (5.57)-(5.58). From the proof of Theorem 5.3.1 (ii) we have that $n^{\beta/2}\nabla L_n^{(\beta)}(\theta_0) \overset{law}{\to} N(0, A(\theta_0))$, 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 (5.17) follows from

$$I_n(\beta) := \mathrm{E}|\nabla\widetilde{L}_n^{(\beta)}(\theta_0) - \nabla L_n^{(\beta)}(\theta_0)| \;=\; o(n^{-\beta/2}). \tag{5.18}$$

We have $I_n(\beta) \le \sup_{n-[n^\beta]\le t\le n} \mathrm{E}|\nabla l_t(\theta_0) - \nabla\widetilde{l}_t(\theta_0)|$ and (5.18) follows from

$$\mathrm{E}|\nabla l_t(\theta_0) - \nabla\widetilde{l}_t(\theta_0)| \;=\; o(t^{-\beta/2}), \qquad t \to \infty. \tag{5.19}$$

Write $\|\xi\|_p := \mathrm{E}^{1/p}|\xi|^p$ for L^p-norm of r.v. ξ. Using $|\nabla(l_t(\theta_0) - \widetilde{l}_t(\theta_0))| \le r_t^2|\nabla(\sigma_t^{-2}(\theta_0) - \widetilde{\sigma}_t^{-2}(\theta_0))| + |\nabla(\log\sigma_t^2(\theta_0) - \log\widetilde{\sigma}_t^2(\theta_0))|$ and assumption $\mathrm{E}|r_t|^5 < \infty$, relation (5.19) follows from

$$\begin{aligned} \|\sigma_t^{-4}\partial_i\sigma_t^2 - \widetilde{\sigma}_t^{-4}\partial_i\widetilde{\sigma}_t^2\|_{5/3} &= O(t^{d_0-1/2}\log t) \quad \text{and} \\ \|\sigma_t^{-2}\partial_i\sigma_t^2 - \widetilde{\sigma}_t^{-2}\partial_i\widetilde{\sigma}_t^2\|_1 &= O(t^{d_0-1/2}\log t), \qquad i = 1, \dots, 5, \end{aligned} \tag{5.20}$$

where $\sigma_t^2 := \sigma_t^2(\theta_0)$, $\widetilde{\sigma}_t^2 := \widetilde{\sigma}_t^2(\theta_0)$, $\partial_i\sigma_t^2 := \partial_i\sigma_t^2(\theta_0)$, $\partial_i\widetilde{\sigma}_t^2 := \partial_i\widetilde{\sigma}_t^2(\theta_0)$. Below, we prove the first relation (5.20) only, 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 $\sigma_t^2 \ge \omega_1^2/(1-\gamma_2) > 0, \widetilde{\sigma}_t^2 \ge \omega_1^2/(1-\gamma_2) > 0$, relation the first relation in (5.20) follows from

$$\begin{aligned} \|(\sigma_t^2 - \widetilde{\sigma}_t^2)(\partial_i\sigma_t^2/\sigma_t)\|_{5/3} &= O(t^{d_0-1/2}) \quad \text{and} && (5.21)\\ \|\partial_i\sigma_t^2 - \partial_i\widetilde{\sigma}_t^2\|_{5/3} &= O(t^{d_0-1/2}\log t), \qquad i = 1,\dots,5. && (5.22)\end{aligned}$$

Consider (5.21). 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 (5.32). Hence, (5.21) follows from

$$\|\sigma_t^2 - \widetilde{\sigma}_t^2\|_{5/2} = O(t^{d_0-1/2}). \qquad (5.23)$$

To show (5.23), similarly as in the proof of (5.43) 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 (4.57), i.e., $U_{t,1} = \sum_{\ell=1}^{t-1}\gamma_0^\ell\Big\{\Big(a_0 + c_0Y_{t-\ell}\Big)^2 - \Big(a_0+c_0\widetilde{Y}_{t-\ell}\Big)^2\Big\}$, $U_{t,2} = \sum_{\ell=t}^{\infty}\gamma_0^\ell\Big\{\omega_0^2+\Big(a_0+c_0Y_{t-\ell}\Big)^2\Big\}$ and $Y_t := Y_t(d_0), \widetilde{Y}_t := \widetilde{Y}_t(d_0)$. We have $|U_{t,1}| \le C\sum_{\ell=1}^{t-1}\gamma_0^\ell|Y_{t-\ell}-\widetilde{Y}_{t-\ell}|(1+|Y_{t-\ell}|+|\widetilde{Y}_{t-\ell}|)$, $|U_{t,2}| \le C\sum_{\ell=t}^{\infty}\gamma_0^\ell(1+|Y_{t-\ell}|^2)$ and hence

$$\begin{aligned}\|\sigma_t^2 - \widetilde{\sigma}_t^2\|_{5/2} &\le C\Big\{\sum_{\ell=1}^{t-1}\gamma_0^\ell\|(Y_{t-\ell} - \widetilde{Y}_{t-\ell})(1 + |Y_{t-\ell}| + |\widetilde{Y}_{t-\ell}|)\|_{5/2}\\ &+ \sum_{\ell=t}^{\infty}\gamma_0^\ell(1 + \|Y_{t-\ell}\|_5)\Big\}\\ &\le C\Big\{\sum_{\ell=1}^{t-1}\gamma_0^\ell\|Y_{t-\ell} - \widetilde{Y}_{t-\ell}\|_5 + \sum_{\ell=t}^{\infty}\gamma_0^\ell\Big\}, \qquad (5.24)\end{aligned}$$

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 (4.13). In a similar way from (4.13) it follows that

$$\|Y_{t-\ell} - \widetilde{Y}_{t-\ell}\|_5 \le C\Big\{\sum_{j>t-\ell} j^{2(d_0-1)}\Big\}^{1/2} \le C(t-\ell)^{d_0-1/2}. \qquad (5.25)$$

Substituting (5.25) into (5.24) we obtain

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

proving (5.23).

It remains to show (5.22). Similarly as above, $\partial_i\sigma_t^2 - \partial_i\widetilde{\sigma}_t^2 = \partial_i U_{t,1} + \partial_i U_{t,2}$, where $\partial_i U_{t,j} := \partial_i U_{t,j}(\theta_0), j = 1, 2$. Then (5.22) follows from

$$\|\partial_i U_{t,1}\|_{5/3} = O(t^{d_0-1/2}\log t) \quad \text{and} \quad \|\partial_i U_{t,2}\|_{5/3} = o(t^{d_0-1/2}), \quad i = 1, \ldots, 5. \quad (5.26)$$

For $i = 1$, the proof of (5.26) is similar to (5.24). Consider (5.26) for $2 \le i \le 5$. Denote $V_t(\theta) := 2a + c(Y_t(d) + \widetilde{Y}_t(d))$, $V_t := V_t(\theta_0), \partial_i V_t := \partial_i V_t(\theta_0)$, 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 (5.25) above. Hence, the first relation in (5.26) follows from (5.25) 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 (5.22), and the proof of the second relation in (5.26) is analogous. This proves (5.19) and completes the proof of Theorem 5.3.2. □

Remark 5.3.1 As noted above, the moment conditions of Theorems 5.3.1 and 5.3.2 are similar to those in Beran and Schützner (2009) for the LARCH model. Particularly, condition (M_5') in Beran and Schützner (2009), Theorems 2 and 5, for asymptotic normality of estimators ensures $\mathrm{E}|r_t|^5 < \infty$. This situation is very different from GARCH models where strong consistency and asymptotic normality of QML estimators holds under virtually no moment assumption on the observed process (see e.g. Francq and Zakoian (2010a), 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 (5.8) 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 (5.8) are much more difficult to control than in the GARCH case, where these quantities are bounded (see Francq and Zakoian (2010a), proof of Theorem 7.2).

Remark 5.3.2 We expect that our results can be extended to more general parametric coefficients, e.g. fractional filters $b_j(c,d), j \ge 1$ with transfer function

$$\sum_{j=1}^{\infty} \mathrm{e}^{-\mathrm{i}jx} b_j(c,d) = g(c,d)((1-\mathrm{e}^{\mathrm{i}x})^{-d} - 1), \quad x \in [-\pi, \pi],$$

where $g(c,d)$ is a smooth function of $(c,d) \in (0,\infty) \times (0,1/2)$. Particularly,

$$b_j(c,d) := g(c,d) \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \sim \frac{g(c,d)}{\Gamma(d)} j^{d-1}, \quad j \to \infty \tag{5.27}$$

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 e.g. Giraitis et al. (2012), Chapter 7. See also Beran and Schützner (2009), Section 2.2. The important condition used in our proofs and satisfied by $b_j(c,d)$ in (5.27) is that the 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, $\partial_d^i b_j^0(d) \sim b_j^0(d)(\log j)^i \sim \Gamma(d)^{-1} j^{d-1} (\log j)^i$, $j \to \infty$, $i = 2,3$.

5.4 Lemmas

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

Lemma 5.4.1 *Let* $\mathrm{E}|r_t|^{2+p} < \infty$*, for some integer* $p \ge 1$*. Then for any* $\boldsymbol{i} \in \mathbb{N}^5$, $0 < |\boldsymbol{i}| \le p$*,*

$$\mathrm{E} \sup_{\theta \in \Theta} |\partial^{\boldsymbol{i}} l_t(\theta)| < \infty. \tag{5.28}$$

Moreover, if $\mathrm{E}|r_t|^{2+p+\epsilon} < \infty$ *for some* $\epsilon > 0$ *and* $p \in \mathbb{N}$ *then for any* $\boldsymbol{i} \in \mathbb{N}^5$, $0 \le |\boldsymbol{i}| \le p$

$$\mathrm{E} \sup_{\theta \in \Theta} |\partial^{\boldsymbol{i}} (l_t(\theta) - \widetilde{l}_t(\theta))| \to 0, \qquad t \to \infty. \tag{5.29}$$

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

$$\partial^{\boldsymbol{i}} \sigma_t^{-2}(\theta) = \sum_{\nu=1}^{|\boldsymbol{i}|}(-1)^\nu \nu!\, \sigma_t^{-2(1+\nu)}(\theta) \sum_{\boldsymbol{j}_1+\cdots+\boldsymbol{j}_\nu=\boldsymbol{i}} \chi_{\boldsymbol{j}_1,\dots,\boldsymbol{j}_\nu} \prod_{k=1}^{\nu} \partial^{\boldsymbol{j}_k}\sigma_t^2(\theta), \quad (5.30)$$

$$\partial^{\boldsymbol{i}} \log \sigma_t^{2}(\theta) = \sum_{\nu=1}^{|\boldsymbol{i}|}(-1)^{\nu-1} (\nu-1)!\, \sigma_t^{-2\nu}(\theta) \sum_{\boldsymbol{j}_1+\cdots+\boldsymbol{j}_\nu=\boldsymbol{i}} \chi_{\boldsymbol{j}_1,\dots,\boldsymbol{j}_\nu} \prod_{k=1}^{\nu} \partial^{\boldsymbol{j}_k}\sigma_t^2(\theta),$$

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 $\boldsymbol{j}_k \neq \mathbf{0}, k = 1,\dots,\nu$, and $\chi_{\boldsymbol{j}_1,\dots,\boldsymbol{j}_\nu}$ is a combinatorial factor depending only on $\boldsymbol{j}_k, 1 \le k \le \nu$.

Let us prove (5.28). We have $|\partial^{\boldsymbol{i}} l_t(\theta)| \le r_t^2 |\partial^{\boldsymbol{i}} \sigma_t^{-2}(\theta)| + |\partial^{\boldsymbol{i}} \log \sigma_t^2(\theta)|$. Hence using (5.30) and the fact that $\sigma_t^2(\theta) \ge \omega^2/(1-\gamma) \ge \omega_1^2/(1-\gamma_2) > 0$ we obtain

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

Therefore by Hölder's inequality

$$\begin{aligned} \mathrm{E}\sup_{\theta\in\Theta} |\partial^{\boldsymbol{i}} l_t(\theta)| \ &\le\ C(\mathrm{E}(r_t^2+1)^{(2+p)/2})^{2/(2+p)} \\ &\quad \times \sum_{\nu=1}^{|\boldsymbol{i}|} \sum_{\boldsymbol{j}_1+\cdots+\boldsymbol{j}_\nu=\boldsymbol{i}} \prod_{k=1}^{\nu} \mathrm{E}^{1/q_k}\Big(\sup_{\theta\in\Theta} |\partial^{\boldsymbol{j}_k} \sigma_t^2(\theta)|/\sigma_t(\theta)\Big)^{q_k}, \end{aligned} \quad (5.31)$$

where $\sum_{j=1}^{\nu} 1/q_j \le p/(2+p)$. Note $|\boldsymbol{i}| = \sum_{k=1}^{\nu} |\boldsymbol{j}_k|$ and hence 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 (5.31) and condition $\mathrm{E}|r_t|^{2+p} \le C$, relation (5.28) follows from

$$\mathrm{E}\sup_{\theta\in\Theta} \Big(|\partial^{\boldsymbol{j}} \sigma_t^2(\theta)|/\sigma_t(\theta)\Big)^{(2+p)/|\boldsymbol{j}|} < \infty \quad (5.32)$$

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

$$\partial_i \sigma_t^2(\theta) \;=\; \begin{cases} \sum_{\ell=1}^\infty \ell\gamma^{\ell-1}\Big\{\omega^2 + \Big(a + cY_{t-\ell}(d)\Big)^2\Big\}, & \theta_i = \gamma, \\ \sum_{\ell=0}^\infty \gamma^\ell 2\omega, & \theta_i = \omega, \\ \sum_{\ell=0}^\infty \gamma^\ell 2\Big(a + cY_{t-\ell}(d)\Big), & \theta_i = a, \\ \sum_{\ell=0}^\infty \gamma^\ell 2\Big(a + cY_{t-\ell}(d)\Big)Y_{t-\ell}(d), & \theta_i = c, \\ \sum_{\ell=0}^\infty \gamma^\ell 2c\Big(a + cY_{t-\ell}(d)\Big)\partial_d Y_{t-\ell}(d), & \theta_i = d. \end{cases} \tag{5.33}$$

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

$$\sup_{\theta\in\Theta}\Big|\frac{\partial_i\sigma_t^2(\theta)}{\sigma_t(\theta)}\Big| \;\le\; C(1 + J_{t,0} + J_{t,1}), \quad i = 1,\dots,5, \quad \text{where} \tag{5.34}$$
$$J_{t,0} \;:=\; \sum_{\ell=0}^\infty \bar\gamma^\ell \sup_{d\in[d_1,d_2]}|Y_{t-\ell}(d)|, \qquad J_{t,1} \;:=\; \sum_{\ell=0}^\infty \bar\gamma^\ell \sup_{d\in[d_1,d_2]}|\partial_d Y_{t-\ell}(d)|.$$

Consider (5.34) 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 $|\partial_\gamma\sigma_t^2(\theta)|/\sigma_t(\theta) \le \Big(\sum_{\ell=1}^\infty \ell^2\gamma^{\ell-2}\Big\{\omega^2 + \Big(a + cY_{t-\ell}(d)\Big)^2\Big\}\Big)^{1/2} \le C(1 + J_{t,0})$ uniformly in $\theta \in \Theta$, proving (5.34) 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 $\theta_i = a$, (5.34) is immediate from (5.33), proving (5.34).

With (5.34) in mind, (5.32) for $|\boldsymbol{j}| = 1$ follows from

$$\mathrm{E}J_{t,i}^{2+p} \;=\; \mathrm{E}\Big(\sum_{\ell=0}^\infty \bar\gamma^\ell \sup_{d\in[d_1,d_2]}|\partial_d^i Y_{t-\ell}(d)|\Big)^{2+p} \;<\; \infty, \quad i = 0, 1. \tag{5.35}$$

Using Minkowski's inequality and stationarity of $\{Y_t(d)\}$ we obtain $\mathrm{E}^{1/(2+p)}J_{t,i}^{2+p} \le \sum_{\ell=0}^\infty \bar\gamma^\ell\ \mathrm{E}^{1/(2+p)}\sup_d|\partial_d^i Y_{t-\ell}(d)|^{2+p} \le C(\mathrm{E}\sup_d|\partial_d^i Y_t(d)|^{2+p})^{1/(2+p)}$, 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 (2009), Lemma 1 (b) and the inequality $xy \le x^q/q + y^{q'}/q', x, y > 0, 1/q + 1/q' = 1$ we obtain

$$\begin{aligned} \sum_{i=0}^1 \mathrm{E}J_{t,i}^{2+p} \;&\le\; C\sum_{i=0}^1 \mathrm{E}\sup_{d\in[d_1,d_2]}|\partial_d^i Y_t(d)|^{2+p} \\ &\le\; C\sum_{i=0}^2 \sup_{d\in[d_1,d_2]}\mathrm{E}|\partial_d^i Y_t(d)|^{2+p} \;<\; \infty \end{aligned} \tag{5.36}$$

since

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

according to condition $\mathrm{E}|r_t|^{2+p} < C$, Rosenthal's inequality in (4.13) 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 (5.32) for $|\boldsymbol{j}| = 1$.

The proof of (5.32) for $2 \le |\boldsymbol{j}| \le p$ is simpler since it reduces to

$$\mathrm{E}\sup_{\theta\in\Theta}|\partial^{\boldsymbol{j}}\sigma_t^2(\theta)|^{(p+2)/2} \;<\; \infty, \qquad 2 \le |\boldsymbol{j}| \le p. \tag{5.37}$$

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 $\sup_{\theta\in\Theta}|\partial^{\boldsymbol{j}}\sigma_t^2(\theta)| \le CJ_{t,0}$ follows as in (5.34) implying (5.37) as in (5.36) above. Next, let $\boldsymbol{j}' \ne \mathbf{0}$. Denote

$$Q_t^2(\theta) := \omega^2 + \Big(a + cY_t(d)\Big)^2 \tag{5.38}$$

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 (5.32) follows from

$$\mathrm{E}\sup_{\theta\in\Theta}|\partial^{\boldsymbol{j}}Q_t^2(\theta)|^{(p+2)/2} \;<\; \infty. \tag{5.39}$$

For $j_2 \ne 0$ (recall $\theta_2 = \omega$) the derivative in (5.39) is trivial so that it suffices to check (5.39) for $j_1 = 0$ only. Then applying Faà di Bruno's rule we get

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

and hence (5.39) reduces to

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

whose proof is similar to (5.35) above. This ends the proof of (5.28).

The proof of (5.29) 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 (5.28) it suffices to show

$$\begin{aligned}
&\mathrm{E}\sup_{\theta\in\Theta}|\partial^{\boldsymbol{i}}(\sigma_t^{-2}(\theta)-\widetilde{\sigma}_t^{-2}(\theta))|^{\frac{p+2}{p}}\to 0 \quad \text{and}\\
&\mathrm{E}\sup_{\theta\in\Theta}|\partial^{\boldsymbol{i}}(\log\sigma_t^{2}(\theta)-\log\widetilde{\sigma}_t^{2}(\theta))|^{\frac{p+2}{p}}\to 0. \qquad (5.40)
\end{aligned}$$

Below, we prove the first relation in (5.40) only, the proof of the second one being analogous.

Using the differentiation rule in (5.30) we have that

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

where

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

Whence, (5.40) follows from

$$\sup_{\theta\in\Theta}|W_t^{\boldsymbol{j}_1,\ldots,\boldsymbol{j}_\nu}(\theta)-\widetilde{W}_t^{\boldsymbol{j}_1,\ldots,\boldsymbol{j}_\nu}(\theta)| \;\to_p\; 0, \qquad t\to\infty \qquad (5.41)$$

and

$$\mathrm{E}\sup_{\theta\in\Theta}\Big(|W_t^{\boldsymbol{j}_1,\ldots,\boldsymbol{j}_\nu}(\theta)|+|\widetilde{W}_t^{\boldsymbol{j}_1,\ldots,\boldsymbol{j}_\nu}(\theta)|\Big)^{(p+2+\epsilon)/p} \;\le\; C<\infty \qquad (5.42)$$

for some constants $\epsilon>0$ and $C>0$ independent of t. In turn, (5.41) and (5.42) follow from

$$\sup_{\theta\in\Theta}|\partial^{\boldsymbol{j}}(\sigma_t^2(\theta)-\widetilde{\sigma}_t^2(\theta))| \;\to_{\mathrm{p}}\; 0, \qquad t\to\infty \qquad (5.43)$$

and

$$\begin{aligned}
&\mathrm{E}\sup_{\theta\in\Theta}\Big(|\partial^{\boldsymbol{j}}\sigma_t^2(\theta)|/\sigma_t(\theta)\Big)^{(2+p+\epsilon)/|\boldsymbol{j}|} \;<\; C, \qquad (5.44)\\
&\mathrm{E}\sup_{\theta\in\Theta}\Big(|\partial^{\boldsymbol{j}}\widetilde{\sigma}_t^2(\theta)|/\widetilde{\sigma}_t(\theta)\Big)^{(2+p+\epsilon)/|\boldsymbol{j}|} \;<\; C,
\end{aligned}$$

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

Using condition $\mathrm{E}|r_t|^{2+p+\epsilon}<C$, relations in (5.44) can be proved analogously to (5.32) and we omit the details. Consider (5.43). Split $\sigma_t^2(\theta)-\widetilde{\sigma}_t^2(\theta)=U_{t,1}(\theta)+U_{t,2}(\theta)$,

where

$$U_{t,1}(\theta) \ := \ \sum_{\ell=1}^{t-1}\gamma^\ell\Big\{\Big(a+cY_{t-\ell}(d)\Big)^2-\Big(a+c\widetilde{Y}_{t-\ell}(d)\Big)^2\Big\}, \tag{5.45}$$

$$U_{t,2}(\theta) \ := \ \sum_{\ell=t}^{\infty}\gamma^\ell\Big\{\omega^2+\Big(a+cY_{t-\ell}(d)\Big)^2\Big\}.$$

Then $\sup_{\theta\in\Theta}|\partial^{\boldsymbol{j}}U_{t,i}(\theta)| \to_{\mathrm{p}} 0,\ t\to\infty,\ i=1,2$ follows by using Assumption (B) and considering the bounds on the derivatives as in the proof of (5.32). For instance, let us prove (5.43) for $\partial^{\boldsymbol{j}} = \partial_d,\ |\boldsymbol{j}| = 1$. We have $|\partial_d U_{t,1}(\theta)| \le C\sum_{\ell=1}^{t-1}\gamma^\ell\big\{(1+|\bar{Y}_{t-\ell}(d)|)|\partial_d(Y_{t-\ell}(d) - \widetilde{Y}_{t-\ell}(d))| + |\partial_d Y_{t-\ell}(d)|\,|Y_{t-\ell}(d) - \widetilde{Y}_{t-\ell}(d)|\big\}$.
Hence, $\sup_{\theta\in\Theta}|\partial_d U_{t,1}(\theta)| \to_{\mathrm{p}} 0$ follows from $0\le\gamma\le\gamma_2<1$ and

$$\mathrm{E}\sup_{d\in[d_1,d_2]}\big(|Y_t(d)-\widetilde{Y}_t(d)|^2+|\partial_d(Y_t(d)-\widetilde{Y}_t(d))|^2\big) \ \to \ 0 \quad \text{and} \tag{5.46}$$

$$\mathrm{E}\sup_{d\in[d_1,d_2]}\big(|Y_t(d)|^2+|\widetilde{Y}_t(d)|^2+|\partial_d Y_t(d)|^2+|\partial_d\widetilde{Y}_t(d)|^2\big) \ \le \ C. \tag{5.47}$$

The proof of (5.47) mimics that of (5.36) and therefore is omitted. To show (5.46), 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 (5.36) to show that the l.h.s. of (5.47) does not exceed $C\sup_{d\in[d_1,d_2]}\sum_{i=0}^2\mathrm{E}|\partial_d^i(Y_t(d)-\widetilde{Y}_t(d))|^2 \le C\sup_{d\in[d_1,d_2]}\sum_{j=t}^\infty j^{2(d-1)}\ (1+\log^2 j) \to 0,\ t\to\infty$. This proves (5.43) for $|\boldsymbol{j}| = 1$ and $\partial^{\boldsymbol{j}} = \partial_d$. The remaining cases in (5.43) follow similarly and we omit the details. This proves (5.29) and completes the proof of Lemma 5.4.1. $\square$

Lemma 5.4.2 *The function $L(\theta), \theta\in\Theta$ in* (5.13) *is bounded and continuous. Moreover, it attains its unique minimum at $\theta=\theta_0$.*

Proof. 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

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

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

$$\sigma_t^2(\theta) = \sigma_t^2(\theta_0) \qquad (\mathrm{P}_{\theta_0}-\text{a.s.}) \tag{5.48}$$

Thus, it remains to show that (5.48) 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. $\xi,\ \mathrm{E}|\xi|<\infty$, where $\mathcal{F}_s = \sigma(\zeta_u, u\le s)$

(see Section 5.2). (5.48) implies

$$0 = P_s(\sigma_t^2(\theta) - \sigma_t^2(\theta_0)) = P_s(Q_t^2(\theta) - Q_t^2(\theta_0)) + (\gamma - \gamma_0) P_s \sigma_{t-1}^2(\theta_0), \quad \forall\, s \le t-1, \tag{5.49}$$

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 (5.38). We have

$$\begin{aligned} P_s Q_t^2(\theta) &= 2ab_{t-s}(\theta) r_s + 2b_{t-s}(\theta) r_s \sum_{u<s} b_{t-u}(\theta) r_u + \sum_{s\le u<t} b_{t-u}^2(\theta) P_s r_u^2 \qquad (5.50)\\ &= 2ab_{t-s}(\theta)\zeta_s \sigma_s(\theta_0) + 2b_{t-s}(\theta)\zeta_s\sigma_s(\theta_0) \sum_{u<s} b_{t-u}(\theta) r_u \\ &+ \sum_{s<u<t} b_{t-u}^2(\theta) P_s \sigma_u^2(\theta_0) + b_{t-s}^2(\theta)(\zeta_s^2 - 1)\sigma_s^2(\theta_0). \end{aligned}$$

Whence and from (5.49) for $s = t-1$ using $P_{t-1}\sigma_{t-1}^2(\theta_0) = 0$ we obtain

$$C_1(\theta, \theta_0)\zeta_{t-1}^2 + 2C_2(\theta,\theta_0)\zeta_{t-1} - C_1(\theta,\theta_0) = 0 \tag{5.51}$$

where

$$\begin{aligned} C_1(\theta,\theta_0) &:= (c^2 - c_0^2)\sigma_{t-1}(\theta_0),\\ C_2(\theta,\theta_0) &:= (ac - a_0c_0) + \sum_{u<t-1} (c^2(t-u)^{d-1} - c_0^2(t-u)^{d_0-1}) r_u. \end{aligned}$$

Since $C_i(\theta,\theta_0), i = 1,2$ are $\mathcal{F}_{t-2}$-measurable, (5.51) implies $C_1(\theta,\theta_0) = C_2(\theta,\theta_0) = 0$, particularly, $c = c_0$ since $\sigma_{t-1}(\theta_0) \ge \omega > 0$. 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 $0 = \mathrm{E}(\sum_{u<t-1}((t-u)^{d-1} - (t-u)^{d_0-1}) r_u)^2 = \mathrm{E} r_0^2 \sum_{j\ge 2}(j^{d-1} - j^{d_0-1})^2 = 0$, or $d = d_0$. Consequently, $P_s(Q_t^2(\theta) - Q_t^2(\theta_0)) = 0$ for any $s \le t-1$ and hence $\gamma = \gamma_0$ in view of (5.49). 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 $\theta = \theta_0$ and the lemma, too. □

Lemma 5.4.3 *Let* $\mathrm{E} r_0^4 < \infty$. *Then matrices* $A(\theta)$ *and* $B(\theta)$ *in* (5.14) *are well-defined and strictly positive definite for any* $\theta \in \Theta$.

Proof. From (5.15), it suffices to show that

$$\nabla \sigma_t^2(\theta)\lambda^T = 0 \tag{5.52}$$

for some $\theta \in \Theta$ and $\lambda \in \mathbb{R}^5, \lambda \ne 0$ leads to a contradiction. To the last end, we use a similar projection argument as in the proof of Lemma 5.4.2. First, note that

$\sigma_t^2(\theta) = Q_t^2(\theta) + \gamma\sigma_{t-1}^2(\theta)$ implies

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

where $\nabla_4 = (\partial/\theta_2, \dots, \partial\theta_5)$. Hence and using the fact that (5.52) holds for any $t \in \mathbb{Z}$ by stationarity, from (5.52) we obtain

$$(\sigma_{t-1}^2(\theta), \nabla_4 Q_t^2(\theta))\lambda^T = 0. \tag{5.53}$$

Thus,

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

c.f. (5.49). For $s = t-1$ using $P_{t-1}\sigma_{t-1}^2(\theta) = 0$, $P_{t-1}\nabla_4 Q_t^2(\theta) = \nabla_4 P_{t-1} Q_t^2(\theta)$ by differentiating (5.50) similarly to (5.51) we obtain

$$D_1(\lambda)\zeta_{t-1}^2 + 2D_2(\lambda)\zeta_{t-1} - D_1(\lambda) = 0 \tag{5.54}$$

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

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

$\lambda = (\lambda_1, \dots, \lambda_5)^T$. As in (5.51), $D_i(\lambda), i = 1,2$ are $\mathcal{F}_{t-2}$-measurable, (5.54) 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 $c \ne 0$ we get $\lambda_3 = 0$ and then $\lambda_4 = 0$ since $\mathrm{E}(\sum_{u<t-1}(t-u)^{d-2}\log(t-u) r_u)^2 \ne 0$. The above facts allow to rewrite (5.53) as $2\omega\lambda_2 + \lambda_1\sigma_{t-1}^2(\theta) = 0$. Unless both λ_1, λ_2 vanish, the last equation means that either $\lambda_1 \ne 0$ and $\{\sigma_t^2(\theta)\}$ is a deterministic process which contradicts $c \ne 0$, or $\lambda_1 = 0, \lambda_2 \ne 0$ and $\omega = 0$, which contradicts $\omega \ne 0$. Lemma 5.4.3 is proved. $\square$

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

Lemma 5.4.4 *(i) Let* $\mathrm{E}|r_t|^3 < \infty$. *Then*

$$\sup_{\theta\in\Theta}|L_n(\theta) - L(\theta)| \;\overset{a.s.}{\to}\; 0 \qquad \textit{and} \qquad \mathrm{E}\sup_{\theta\in\Theta}|L_n(\theta) - \widetilde{L}_n(\theta)| \;\to\; 0. \tag{5.55}$$

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

$$\sup_{\theta\in\Theta}|\nabla L_n(\theta) - \nabla L(\theta)| \;\overset{a.s.}{\to}\; 0 \qquad \textit{and} \qquad \mathrm{E}\sup_{\theta\in\Theta}|\nabla L_n(\theta) - \nabla\widetilde{L}_n(\theta)| \;\to\; 0. \tag{5.56}$$

(iii) Let $\mathrm{E}|r_t|^5 < \infty$. *Then* $\nabla^T\nabla L(\theta) = \mathrm{E}\nabla^T\nabla \ell_t(\theta) = B(\theta)$ *(see* (5.14)*) and*

$$\sup_{\theta\in\Theta} |\nabla^T\nabla L_n(\theta) - \nabla^T\nabla L(\theta)| \overset{a.s.}{\to} 0, \tag{5.57}$$

$$\mathrm{E}\sup_{\theta\in\Theta} |\nabla^T\nabla L_n(\theta) - \nabla^T\nabla \widetilde{L}_n(\theta)| \to 0. \tag{5.58}$$

Proof. Consider the first relation in (5.55). The pointwise convergence $L_n(\theta) \overset{a.s.}{\to} L(\theta)$ follows by ergodicity of $\{l_t(\theta)\}$ and the uniform convergence in (5.55) from $\mathrm{E}\sup_{\theta\in\Theta}|\nabla l_t(\theta)| < \infty$, c.f. Beran and Schützner (2009), proof of Lemma 3, which in turn follows from of Lemma 5.4.1 (5.28) with $p = 1$. The proof of the second relation in (5.55) is immediate from Lemma 5.4.1 (5.29) with $p = 0, \epsilon = 1$. The proof of the statements (ii) and (iii) using Lemma 5.4.1 is similar and is omitted. □

5.5 Simulation study

In this section we present a short simulation study of the performance of the QMLE for the GQARCH model in (5.4). The GQARCH model in (5.4) 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 (5.1) with zero initial condition $\sigma_{-m} = 0$. The numerical optimization procedure minimized the QML function:

$$\widetilde{L}_m = \frac{1}{m}\sum_{t=1}^{m}\left(\frac{r_t^2}{\sigma_t^2} + \log\sigma_t^2\right), \tag{5.59}$$

with

$$r_t = \zeta_t\sigma_t, \quad \sigma_t^2 = \omega^2 + \Big(a + c\sum_{j=1}^{t+m-1} j^{d-1}r_{t-j}\Big)^2 + \gamma\sigma_{t-1}^2, \quad t = 1,\dots,m. \tag{5.60}$$

The QML function in (5.59) can be viewed as a "realistic proxy" to the QML function $\widetilde{L}_n(\theta)$ in (5.12) with $m = n^\beta$ since (5.59)-(5.60) similarly to (5.12) use "auxiliary" observations in addition to $r_1,\dots,r_m$ for computation of m likelihoods in (5.59). However, the number of "auxiliary" observations in (5.59) equals m and does not grow as $m^{1/\beta} = n, 0 < \beta < 1-2d < 1$ in the case of (5.60) and Theorem 5.3.2 (ii), which is completely unrealistic. Despite of the violation of the condition $m = n^\beta$ of Theorem 5.3.2 (ii) in our simulation study, the differences between the sample RMSEs and the theoretical standard deviations are not vital (and sometimes even insignificant), see Table 5.1 below.

Finite-sample performance of the QML estimator $\widetilde{\theta}_m$ minimizing (5.59) 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 above choice of $\theta_0 = (\gamma_0, \omega_0, a_0, d_0, c_0)$ can be explained by the observation that the QML estimation of γ_0, a_0, c_0 appears to be more accurate and stable in comparison with 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 less than 0.05.

The numerical QML minimization was performed 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 optimization procedure is chosen in a way to guarantee Assumption (B) (v) with appropriate $0 < c_i(d, \gamma), i = 1, 2$.

The results of the simulation experiment are presented in Table 5.1 below, which shows the sample R(oot)MSEs 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 above choices of $\theta_0 = (\gamma_0, \omega_0, a_0, d_0, c_0)$. The sample RMSEs in Table 5.1 are confronted with standard deviations (in brackets) of the infinite past estimator in (5.6) computed according to Theorem 5.3.1 (ii) with $\Sigma(\theta_0)$ obtained by inverting a simulated matrix $B(\theta_0)/\kappa_4$.

A general impression from Table 5.1 is that theoretical standard deviations (bracketed entries) are generally smaller than the sample RMSEs, however, these differences become less pronounced with 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 5.1 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 5.1 appear to be more stable and less dependent on θ_0 compared to the bracketed entries (particularly, this applies to errors of $\widetilde{\omega}_m$ and $\widetilde{d}_m$).

		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)

Table 5.1: Sample RMSE of finite past QML estimates $\widetilde{\theta}_m$ in (5.59) of $\theta_0 = (d_0, \omega_0, a_0, c_0, \gamma_0)$ of the GQARCH process in (5.4) 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 brackets stand for asymptotic standard deviations of the estimator $\widetilde{\theta}_n^{(\beta)}, n^\beta = m$ following Theorem 5.3.1 (ii).

Conclusions

The principal goal of this thesis was to introduce *new nonlinear models with long memory* which can be used for modelling of financial returns and statistical inference. Apart from long memory, these models are capable to exhibit other stylized facts such as asymmetry and leverage. The processes studied in the thesis are defined as stationary solutions of certain *nonlinear stochastic difference equations* involving a given i.i.d. “noise”. Apart from solvability issues of these equations which are not trivial by itself, we proved that their solutions exhibit long memory properties as in (1.5) and (1.9). Finally, for a particularly tractable nonlinear parametric model with long memory (GQARCH) we prove consistency and asymptotic normality of quasi-ML estimators.

Appendix A

Lemmas

The proofs of Proposition 4.3.1 and Theorems 4.4.2, 4.6.3 use the following lemmas.

Lemma A.1 *For* $\alpha_j \geq 0, j = 1, 2, \ldots,$ *denote*

$$A_k \;:=\; \alpha_k + \sum_{0<p<k} \sum_{0<i_1<\cdots<i_p<k} \alpha_{i_1}\alpha_{i_2-i_1}\ldots\alpha_{i_p-i_{p-1}}\alpha_{k-i_p}, \qquad k = 1, 2, \ldots. \tag{A.1}$$

Assume that $A := \sum_{j=1}^{\infty} \alpha_j < 1$ *and*

$$\alpha_j \;\leq\; c\, j^{-\gamma} \qquad (\exists\, c > 0,\; \gamma > 1). \tag{A.2}$$

Then there exists $C > 0$ *such that for any* $k \geq 1$

$$A_k \;\leq\; Ck^{-\gamma}. \tag{A.3}$$

Proof. We have $A_k = \sum_{0 \leq p < k} A_{k,p}$, where

$$A_{k,p} := \sum_{0<i_1<\cdots<i_p<k} \alpha_{i_1}\alpha_{i_2-i_1}\ldots\alpha_{i_p-i_{p-1}}\alpha_{k-i_p} \quad (p \geq 1), \quad A_{k,0} := \alpha_k$$

is the inner sum in (A.1). W.l.g., assume $c \geq 1$ in (A.2). Let us prove that there exists $\lambda > 0$ such that

$$A_{k,p} \;\leq\; c(p+2)^{\lambda} A^{p+1} k^{-\gamma}, \qquad \forall\, 0 \leq p < k < \infty. \tag{A.4}$$

Since $A < 1$, so (A.4) and $\sum_{p>0}(p+2)^{\lambda}A^{p+1} < \infty$ together imply (A.3).

By dividing both sides of (A.4) by A^{p+1}, it suffices to show (A.4) for $A = 1$. The proof uses induction on p. Clearly, (A.4) holds for $p = 0$. To prove the induction step $p-1 \to p \ge 1$, note

$$A_{k,p} \;=\; \sum_{0<i<k} \alpha_i A_{k-i,p-1} = \sum_{\frac{k}{p+1}<i<k} \alpha_i A_{k-i,p-1} + \sum_{k-\frac{k}{p+1}\le k-i<k} \alpha_i A_{k-i,p-1}. \quad \text{(A.5)}$$

Here, $\alpha_i \mathbf{1}(i > \frac{k}{p+1}) \le c i^{-\gamma}\mathbf{1}(i > \frac{k}{p+1}) \le c(p+1)^\gamma k^{-\gamma}$ and, similarly, by the inductive assumption

$$A_{k-i,p-1}\mathbf{1}(k-i \ge k - \frac{k}{p+1}) \le c(p+1)^\lambda (k - \frac{k}{p+1})^{-\gamma} = c(p+1)^\lambda \Big(\frac{p+1}{p}\Big)^\gamma k^{-\gamma}.$$

Assumption $A=1$ implies $\sum_{k>0} A_{k,p} = 1$ for any $p \ge 0$. Using the above facts from (A.5) we obtain

$$\begin{aligned} A_{k,p} \;&=\; \frac{c(p+1)^\gamma}{k^\gamma} \sum_{k/(p+1)<i<k} A_{k-i,p-1} + \frac{c(p+1)^\lambda}{k^\gamma}\Big(\frac{p+1}{p}\Big)^\gamma \sum_{k-k/(p+1)\le k-i<k} \alpha_i \\ &\le\; c\Big((p+1)^\gamma + (p+1)^\lambda \Big(\frac{p+1}{p}\Big)^\gamma\Big) k^{-\gamma}. \end{aligned}$$

Hence the proof of the induction step $p-1 \to p \ge 1$ amounts to verifying the inequality $(p+1)^\gamma + (p+1)^\lambda\big(\frac{p+1}{p}\big)^\gamma \le (p+2)^\lambda$, or

$$n^\gamma + n^\lambda \Big(\frac{n}{n-1}\Big)^\gamma \;\le\; (n+1)^\lambda, \qquad n = 2,3,\ldots. \quad \text{(A.6)}$$

The above inequality holds with $\lambda = 3\gamma$. Indeed,

$$\begin{aligned} n^\gamma + n^\lambda \Big(\frac{n}{n-1}\Big)^\gamma \;&=\; n^\lambda (n^{-2\gamma} + \Big(\frac{n}{n-1}\Big)^\gamma) \le n^\lambda (n^{-2} + \Big(\frac{n}{n-1}\Big))^\gamma \\ &\le\; n^\lambda (1 + \frac{1}{n-1} + \frac{1}{n^2})^\gamma \le n^\lambda (1 + \frac{3}{n} + \frac{3}{n^2} + \frac{1}{n^3})^\gamma = (n+1)^\lambda, \end{aligned}$$

proving (A.6) and the lemma, too. $\square$

Lemma A.2 *Assume that* $0 \le \beta < 1$ *and* $\alpha_j \sim c j^{-\gamma}$ $(\exists\, \gamma > 0,\, c > 0)$. *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 $j \ge 0$. $\square$

Appendix B

Nested Volterra series

First we introduce some notation. Let $T \subset \mathbb{Z}$ be a set of integers which is bounded from above, i.e., $\sup\{s : s \in T\} < \infty$. Let $\mathcal{S}_T$ be a class of nonempty subsets $S = \{s_1, \dots, s_n\} \subset T$, $s_1 < \cdots < s_n$. Write $|S|$ for the cardinality of $S \subset Z$. For any $S = \{s_1, \dots, s_n\} \in \mathcal{S}_T$, $S' = \{s'_1, \dots, s'_m\} \in \mathcal{S}_T$, the notation $S \prec S'$ means that $m = n + 1$ and $s_1 = s'_1, \dots, s_n = s'_n < s'_{n+1} = s'_m$. In particular, $S \prec S'$ implies $S \subset S'$ and $|S' \backslash S| = 1$. Note that $\prec$ is not a partial order in $\mathcal{S}_T$ since $S \prec S'$, $S' \prec S''$ do not imply $S \prec S''$. A set $S \in \mathcal{S}_T$ is said *maximal* if there is no $S' \in \mathcal{S}_T$ such that $S \prec S'$. Let $\mathcal{S}_T^{\max}$ denote the class of all maximal elements of $\mathcal{S}_T$.

Definition B.1 *Let $T \subset \mathbb{Z}$ be a set bounded from above, and $\mathcal{S}_T$ be a class of subsets of T. Let $\mathcal{G}_T := \{G_S, S \in \mathcal{S}_T\}$ be a family of measurable functions $G_S = G_{s_1,\dots,s_m} : \mathbb{R} \to \mathbb{R}$ indexed by sets $S = \{s_1, \dots, s_m\} \in \mathcal{S}_T$ and such that $G_S =: a_S$ is a constant function for any maximal set $S \in \mathcal{S}_T^{\max}$. A nested Volterra series is a sum*

$$\begin{aligned} V(\mathcal{G}_T) \;=\; & \sum_{S_1 \in \mathcal{S}_T : |S|=1} \zeta_{S_1} G_{S_1}\Big(\sum_{S_1 \prec S_2} \zeta_{S_2 \backslash S_1} G_{S_2}\Big(\dots \\ & \zeta_{S_{p-1} \backslash S_{p-2}} G_{S_{p-1}}\Big(\sum_{S_{p-1} \prec S_p} \zeta_{S_p \backslash S_{p-1}} G_{S_p}\Big)\Big)\Big), \qquad \text{(B.1)} \end{aligned}$$

where the nested summation is taken over all sequences $S_1 \prec S_1 \prec \cdots \prec S_p \in \mathcal{S}_T^{\max}, p = 1, 2, \dots,$ with the convention that $G_S = a_S$, $S \in \mathcal{S}_T^{\max}$, and $\zeta_S := \zeta_s$ for $S = \{s\}, |S| = 1$.

In particular, when $\mathcal{S}_T = \{S : S \subset T\}$ is the class of all subsets of T, (B.1) can *be rewritten as*

$$V(\mathcal{G}_T) \;=\; \sum_{s_1 \in T} \zeta_{s_1} G_{s_1}\Big(\sum_{s_1 < s_2 \in T} \zeta_{s_2} G_{s_1,s_2}\Big(\dots \\ \zeta_{s_{p-1}} G_{s_1,\dots,s_{p-1}}\Big(\sum_{s_{p-1} < s_p \in T} \zeta_{s_p} G_{s_1,\dots,s_p}\Big)\Big)\Big), \tag{B.2}$$

where the last sum is taken over all maximal sets $\{s_1, \dots, s_p\} \in \mathcal{S}_T^{\max}$.

The following example clarifies the above definition and its relation to the usual Volterra series (Dedecker et al., 2007, p.22).

Example B.1 *Let* $T(t) = (-\infty, t] \cap \mathbb{Z}$, $t \in \mathbb{Z}$ *and* $\mathcal{S}_{T(t)}$ *be the class of all subsets* $S = \{s_1, \dots, s_k\} \subset T(t)$ *having* k *points. Let* $\mathcal{G}_{T(t)} = \{G_S, S \in \mathcal{S}_{T(t)}\}$ *be a family of linear functions*

$$G_S(x) := \begin{cases} x, & S \in \mathcal{S}_T, S \notin \mathcal{S}_{T(t)}^{\max}, \\ a_S = a_{s_1,\dots,s_k}, & S = \{s_1, \dots, s_k\} \in \mathcal{S}_{T(t)}^{\max}. \end{cases}$$

Then

$$V(\mathcal{G}_{T(t)}) \;=\; \sum_{s_1 < \dots < s_k \le t} a_{s_1,\dots,s_k} \zeta_{s_1} \zeta_{s_2} \dots \zeta_{s_k} \;=\; \sum_{S \subset T, |S| = k} a_S \zeta^S, \tag{B.3}$$

$\zeta^S := \zeta_{s_1}\zeta_{s_2}\dots\zeta_{s_k}$, *is the (usual) Volterra series of order* k. *The series in (B.3) converges in mean square if and only if*

$$A_{T(t)} \;:=\; \sum_{s_1 < \dots < s_k \le t} a^2_{s_1,\dots,s_k} < \infty, \tag{B.4}$$

in which case $\mathrm{E}V^2(\mathcal{G}_{T(t)}) = A_{T(t)}$, $\mathrm{E}V(\mathcal{G}_{T(t)}) = 0$.

Proposition B.1 *Let* $T(t) := (-\infty, t] \cap \mathbb{Z}$, $t \in \mathbb{Z}$ *as in Example B.1. Assume that the system* $\mathcal{G}_{T(t)} = \{G_S, S \in \mathcal{S}_{T(t)}\}$ *in Definition B.1 satisfies the following condition*

$$|G_S(x)|^2 \;\le\; \begin{cases} \alpha_S^2 + \beta_S^2 x^2, & S \in \mathcal{S}_{T(t)},\, S \notin \mathcal{S}_{T(t)}^{\max}, \\ \alpha_S^2 (= a_S^2), & S \in \mathcal{S}_{T(t)}^{\max}, \end{cases} \tag{B.5}$$

where α_S, β_S *are real numbers satisfying*

$$\mathcal{A}_{T(t)} \;:=\; \sum_{p \ge 1} \sum_{S_1 \prec S_2 \prec \dots \prec S_p} \beta_{S_1}^2 \beta_{S_2}^2 \dots \beta_{S_{p-1}}^2 \alpha_{S_p}^2 \;<\; \infty, \tag{B.6}$$

where the inner sums are taken over all sequences $S_1 \prec S_2 \prec \cdots \prec S_p$, $S_i \in \mathcal{S}_{T(t)}$, $1 \le i \le p$ *with* $|S_1| = 1$ *and* $S_p \in \mathcal{S}^{\max}_{T(t)}$.

Then, the nested Volterra series $V(\mathcal{G}_{T(t)})$ *in (B.2) converges in mean square and satisfies* $\mathrm{E}V(\mathcal{G}_{T(t)})^2 \le \mathcal{A}_{T(t)}$, $\mathrm{E}V(\mathcal{G}_{T(t)}) = 0$. *Moreover,* $X_t := V(\mathcal{G}_{T(t)})$ *is a projective process with zero mean and coefficients*

$$g_{s,t} \;:=\; G_{S_1}\Big(\sum_{S_1 \prec S_2} \zeta_{S_2\backslash S_1} G_{S_2}\Big(\dots \zeta_{S_{p-1}\backslash S_{p-2}} G_{S_{p-1}}\Big(\sum_{S_{p-1}\prec S_p} \zeta_{S_p \backslash S_{p-1}} G_{S_p}\Big)\Big)\Big) \quad \text{(B.7)}$$

if $S_1 = \{s\} \in \mathcal{S}_{T(t)}$, $g_{s,t} := 0$ *otherwise, where the nested summation is defined as in (B.1).*

Proof. Clearly, the coefficients $g_{s,t}$ in (B.7) satisfy the measurability condition (i) of Definition 3.2.1. Condition (ii) for these coefficients follows by recurrent application of (B.5):

$$\begin{aligned}
\sum_{s \le t} \mathrm{E} g^2_{s,t} &= \sum_{S_1 \in \mathcal{S}_{T(t)} : |S_1| = 1} \mathrm{E} G^2_{S_1}\Big(\sum_{S_1 \prec S_2} \zeta_{S_2\backslash S_1} G_{S_2}(\dots)\Big) \\
&\le \sum_{S_1 \in \mathcal{S}_{T(t)} : |S_1| = 1} \Big(\alpha^2_{S_1} + \beta^2_{S_1} \mathrm{E}\Big(\sum_{S_1 \prec S_2} \zeta_{S_2\backslash S_1} G_{S_2}(\dots)\Big)^2\Big) \\
&\le \sum_{S_1 \in \mathcal{S}_{T(t)} : |S_1| = 1} \Big(\alpha^2_{S_1} + \beta^2_{S_1} \sum_{S_1 \prec S_2} \Big(\alpha^2_{S_2} + \beta^2_{S_2} \mathrm{E}\Big(\sum_{S_2 \prec S_3} \zeta_{S_3\backslash S_2} G_{S_3}(\dots)\Big)^2\Big)\Big) \\
&\le \sum_{S_1 \in \mathcal{S}_{T(t)} : |S_1| = 1} \alpha^2_{S_1}\Big(1 + \beta^2_{S_1} \sum_{S_1 \prec S_2} \alpha^2_{S_2} + \beta^2_{S_1} \sum_{S_1 \prec S_2 \prec S_3} \beta^2_{S_2} \alpha^2_{S_3} + \dots\Big) \\
&= \sum_{p \ge 1} \sum_{S_1 \prec S_2 \prec \cdots \prec S_p} \beta^2_{S_1} \beta^2_{S_2} \dots \beta^2_{S_{p-1}} \alpha^2_{S_p} = \mathcal{A}_{T(t)} < \infty.
\end{aligned}$$

Therefore, $X_t = \sum_{s \le t} g_{s,t} \zeta_s$ is a well-defined projective process and $X_t = V(\mathcal{G}_{T(t)})$. Proposition B.1 is proved. □

Remark B.1 In the case of a usual Volterra series in (B.3), condition (B.5) is satisfied with $\alpha_S = 0, \beta_S = 1$ for $S \in \mathcal{S}_{T(t)}$, $S \notin \mathcal{S}^{\max}_{T(t)}$, and the sums $\mathcal{A}_{T(t)}$ of (B.6) and $A_{T(t)}$ of (B.4) coincide: $\mathcal{A}_{T(t)} = A_{T(t)}$. This fact confirms that condition (B.6) for the convergence of nested Volterra series cannot be generally improved.

Bibliography

Abadir, K., Distaso, W., Giraitis, L., and Koul, H. (2014). Asymptotic normality for weighted sums of linear processes. *Econometric Th.*, 30:252–284.

Baillie, T. and Kapetanios, G. (2008). Nonlinear models for strongly dependent processes with financial applications. *J. Econometrics*, 147:60–71.

Beran, J. (1997). *Statistics for Long Memory Processes*, volume 61 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, New York.

Beran, J., Feng, Y., Gosh, S., and Kulik, R. (2013). *Long Memory Processes: Probabilistic Properties and Statistical Methods.* Springer, Heidelberg.

Beran, J. and Schützner, M. (2009). On approximate pseudo-maximum likelihood estimation for LARCH-processes. *Bernoulli*, 15:1057–1081.

Berkes, I. and Horváth, L. (2003). Asymptotic results for long memory LARCH sequences. *Ann. Appl. Probab.*, 13:641–668.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Black, F. (1976). Studies in stock price volatility changes. In *1976 Business Meeting of the Business and Economics Statist. Sec.*, pages 177–181. Amer. Statist. Assoc.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econometrics*, 31:307–327.

Bollerslev, T., Chou, R., and Kroner, K. (1992). ARCH modeling in finance. A review of the theory and empirical evidence. *J. Econometrics*, 52:5–59.

Bollerslev, T. and Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73:151–184.

Breidt, F. J., Crato, N., and de Lima, P. (1998). On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83:325–348.

Burkholder, D. L. (1973). Distribution functions inequalities for martingales. *Ann. Probab.*, 1:19–42.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, Chichester.

Davis, R. A. and Mikosch, T. (2009). Probabilistic Properties of Stochastic Volatility Models. In Andersen, T., Davis, R., Kreiss, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*, pages 255–268. Springer-Verlag.

Davydov, Y. (1970). The invariance principle for stationary process. *Theory Probab. Appl.*, 15:145–180.

Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, S., and Prieur, C. (2007). *Weak Dependence: With Examples and Applications*, volume 190 of *Lecture Notes in Statistics*. Springer, New York.

Dedecker, J. and Merlevède, F. (2003). The conditional central limit theorem in Hilbert spaces. *Stoch. Process. Appl.*, 108:229–262.

Dedecker, J. and Prieur, C. (2004). Coupling for $\tau-$dependent sequences and applications. *J. Theoretical Probab.*, 17:861–885.

Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Th. Relat. Fields.*, 132:203–236.

Dedecker, J. and Prieur, C. (2007). An empirical central limit theorem for dependent sequences. *Stoch. Process. Appl.*, 117:121–142.

Doukhan, P., Grublytė, I., and Surgailis, D. (2016). A nonlinear model for long memory conditional heteroscedasticity. *Lithuanian Math. J.*, 56(2):164–188.

Doukhan, P., Lang, G., and Surgailis, D. (2012). A class of Bernoulli shifts with long memory: asymptotics of the partial sums process. Preprint.

Doukhan, P., Oppenheim, G., and Taqqu, M. (2003). *Theory and Applications of Long-Range Dependence*. Birkhäuser, Boston.

Engle, R. F. (1982). Autoregressive conditional heterosckedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1008.

Engle, R. F. (1990). Stock volatility and the crash of '87. Discussion. *Rev. Financial Studies*, 3:103–106.

Francq, C. and Zakoian, J.-M. (2009). A tour in the asymptotic theory of GARCH estimation. In Andersen, T., Davis, R., Kreiss, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*, pages 85–111. Springer-Verlag.

Francq, C. and Zakoian, J.-M. (2010a). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, New York.

Francq, C. and Zakoian, J.-M. (2010b). Inconsistency of the MLE and inference based on weighted LS for LARCH models. *J. Econometrics*, 159:151–165.

Giraitis, L., Koul, H. L., and Surgailis, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press, London.

Giraitis, L., Leipus, R., Robinson, P. M., and Surgailis, D. (2004). LARCH, leverage and long memory. *J. Financial Econometrics*, 2:177–210.

Giraitis, L., Leipus, R., and Surgailis, D. (2009). ARCH(∞) models and long memory properties. In Andersen, T., Davis, R., Kreiss, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*, pages 71–84. Springer-Verlag.

Giraitis, L., Robinson, P. M., and Surgailis, D. (2000). A model for long memory conditional heteroskedasticity. *Ann. Appl. Probab.*, 10:1002–1024.

Giraitis, L. and Surgailis, D. (2002). ARCH-type bilinear models with double long memory. *Stoch. Process. Appl.*, 100:275–300.

Giraitis, L., Surgailis, D., and Škarnulis, A. (2016). Integrated $ARCH(\infty)$ processes with finite variance. *preprint*.

Grublytė, I. and Škarnulis, A. (2017). A generalized nonlinear model for long memory conditional heteroscedasticity. *Statistics*, 51(1):123–140.

Grublytė, I. and Surgailis, D. (2014). Projective stochastic equations and nonlinear long memory. *Adv. Appl. Probab.*, 46(4):1–22.

Grublytė, I., Surgailis, D., and Škarnulis, A. (2017). QMLE for quadratic ARCH model with long memory. *Journal of Time Series Analysis*, 38(4):535–551.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Applications*. Academic Press, New York.

Harvey, A. (1998). Long memory in stochastic volatility. In Knight, J. and Satchell, S., editors, *Forecasting Volatility in the Financial Markets*, pages 307–320. Butterworth and Heineman.

Hitchenko, P. (1990). Best constants in martingale version of Rosenthal's inequality. *Ann. Probab.*, 18:1656–1668.

Ho, H.-C. and Hsing, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.*, 25:1636–1669.

Hurst, H. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–808.

Hurvich, C. M. and Soulier, P. (2009). Stochastic Volatility Models with Long Memory. In Andersen, T., Davis, R., Kreiss, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*, pages 345–354. Springer-Verlag.

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

Lamperti, J. W. (1962). Semi-stable stochastic processes. *Trans. Amer. Math. Soc.*, 104:62–78.

Levine, M., Torres, S., and Viens, F. (2009). Estimation for the long-memory parameter in LARCH models, and fractional Brownian motion. *Stat. Inf. Stoch. Process.*, 12:221–250.

Lindner, A. M. (2009). Stationarity, mixing, distributional properties and moments of GARCH$(p, q)-$processes. In Andersen, T., Davis, R., Kreiss, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*, pages 43–69. Springer-Verlag.

Mandelbrot, B. (1965). Une classe de processus stochastiques homothetiques a soi; application a loi climatologique de H. E. Hurst. *Comptes Rendus Academic Sciences Paris*, 240:3274–3277.

Mandelbrot, B. and Van Ness, J. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10:422–437.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59:347–370.

Osękowski, A. (2012). A note on Burkholder-Rosenthal inequality. *Bull. Polish Acad. Sci. Mathematics*, 60:177–185.

Philippe, A., Surgailis, D., and Viano, M.-C. (2006). Invariance principle for a class of non stationary processes with long memory. *C. R. Acad. Sci. Paris*, pages 269–274.

Philippe, A., Surgailis, D., and Viano, M.-C. (2008). Time-varying fractionally integrated processes with nonstationary long memory. *Th. Probab. Appl.*, 52:651–673.

Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Econometrics*, 47:67–84.

Robinson, P. M. (2001). The memory of stochastic volatility models. *J. Econometrics*, 101:195–218.

Robinson, P. M. and Zaffaroni, P. (2006). Pseudo-maximum likelihood estimation of ARCH(∞) models. *The Annals of Statistics*, 34:1049–1074.

Rosenthal, H. P. (1970). On the subspaces of L^p $(p > 2)$ spanned by the sequences of independent random variables. *Israel J. Math.*, 8:273–303.

Samorodnitsky, G. (2007). Long range dependence. *Stochastic Systems*, 1:163—-257.

Sentana, E. (1995). Quadratic ARCH models. *Rev. Econom. Stud.*, 3:77–102.

Shao, X. and Wu, W. (2006). Invariance principles for fractionally integrated nonlinear processes. Preprint http://galton.uchicago.edu/ wbwu/papers/fipfeb3.pdf.

Shephard, N. and Andersen, T. (2009). Stochastic volatility: origins and overview. In Andersen, T., Davis, R., Kreiss, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*, pages 233–254. Springer-Verlag.

Stout, W. (1974). *Almost sure convergence*. Academic Press, New York.

Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models*. Springer, Berlin Heidelberg.

Surgailis, D. (2008). A Quadratic ARCH(∞) model with long memory and Lévy stable behavior of squares. *Adv. Appl. Probab.*, 40:1198–1222.

Taqqu, M. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Zeit. Wahrsch. verw. Geb.*, 50:53–83.

Truquet, L. (2014). On a family of contrasts for parametric inference in degenerate ARCH models. *Econometric Th.*, 30:1165–1206.

von Bahr, B. and Esséen, C.-G. (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Ann. Math. Statist.*, 36:299–303.

Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv för Matematik*, 2:423–443.

Wu, W. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci.*, 102:14150–14154.

Wu, W. and Min, W. (2005). On linear processes with dependent innovations. *Stoch. Process. Appl.*, 115:939–958.