



Munich Personal RePEc Archive

Functional generalized autoregressive conditional heteroskedasticity

Aue, Alexander and Horvath, Lajos and Pellatt, Daniel

University of California Davis, University of Utah, University of
Utah

20 August 2015

Online at <https://mpra.ub.uni-muenchen.de/67702/>
MPRA Paper No. 67702, posted 09 Nov 2015 05:30 UTC

Functional generalized autoregressive conditional heteroskedasticity*

Alexander Aue[†]

Lajos Horváth[‡]

Daniel F. Pellatt[‡]

September 12, 2015

Abstract

Heteroskedasticity is a common feature of financial time series and is commonly addressed in the model building process through the use of ARCH and GARCH processes. More recently multivariate variants of these processes have been in the focus of research with attention given to methods seeking an efficient and economic estimation of a large number of model parameters. Due to the need for estimation of many parameters, however, these models may not be suitable for modeling now prevalent high-frequency volatility data. One potentially useful way to bypass these issues is to take a functional approach. In this paper, theory is developed for a new functional version of the generalized autoregressive conditionally heteroskedastic process, termed fGARCH. The main results are concerned with the structure of the fGARCH(1,1) process, providing criteria for the existence of a strictly stationary solutions both in the space of square-integrable and continuous functions. An estimation procedure is introduced and its consistency verified. A small empirical study highlights potential applications to intraday volatility estimation.

Keywords: Econometrics; Financial time series; Functional data; GARCH processes; Stationary solutions

MSC 2010: Primary: 62M10, Secondary: 62P20, 91B84

1 Introduction

Modeling volatility is one of the prime objectives of financial time series analysis. Research in the area has increased rapidly since the seminal contributions by Engle (1982), who introduced the autoregressive conditional heteroskedastic, ARCH, model, and Bollerslev (1986), who introduced the generalized ARCH, GARCH, model to deal with non-constant and randomly changing volatilities for univariate time series. Methodology based on these processes and their modifications have become major theoretical and applied tools to analyze stock returns, exchange rates and more. The book Francq and Zakoian (2010) provides an excellent introduction to volatility processes and their applications. The interested reader may find detailed

*This research was partially supported by NSF grants DMS 1209226, DMS 1305858 and DMS 1407530

[†]Department of Statistics, University of California, Davis, CA 95616, USA, email: aaue@ucdavis.edu

[‡]Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA, emails: horvath@math.utah.edu, daniel.pellatt@gmail.com

accounts on theoretical properties such as conditions for the existence of stationary solutions and finiteness of moments as well as methodology for the estimation of model parameters.

With ever more data being collected in today's financial sector, there has come a need to address new phenomena observed when dealing with high-frequency, intra-day observations. While multivariate volatility processes such as the ones of Engle (2002) and Engle and Kroner (1995) might be used to model high-frequency returns, these models are typically associated with the difficult task of estimating a large number of parameters (see Chapter 11 of Francq and Zakoian, 2010). Surveys of multivariate GARCH models and related volatility processes may be found in Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2009). To alleviate the numerical burden of the parameter estimation, the variance targeting method of Engle and Mezrich (1996) can be used, which was recently extended to cover multivariate GARCH models by Francq et al. (2011, 2015). These methods, however, may still fail to produce satisfactory results in a high-frequency setting.

Modeling data as a collection of functions was popularized through the work of Ramsay and Silvermann (2005). While this approach has now started to become relevant for the analysis of high-frequency volatility data, much of the research in the area has so far been devoted to homoskedastic functional time series. The books by Bosq (2010) and Horváth and Kokoszka (2012) provide an account of the state-of-the-art research on dependent sequences of random functions, including functional autoregressive processes. More recently, prediction of functional time series was considered in Aue et al. (2015), while estimation in dependent functional linear models was discussed in Aue et al. (2014).

A first step to introduce a functional heteroskedastic framework in the tradition of Engle (1982) was undertaken in Hörmann et al. (2013). These authors found conditions for the existence of functional ARCH(1) processes, for which the conditional variance depends on the whole (intra-day) path of the previous observation. In the spirit of Bollerslev (1986), this paper introduces a functional model in which the conditional volatility function of the present observation is given as a functional linear combination of the (intra-day) paths of both the past squared observation and its volatility function. The following definition is central.

Definition 1.1. *A sequence of random functions $(y_i: i \in \mathbb{Z})$ is called a functional GARCH process of orders $(1,1)$, abbreviated by fGARCH(1,1), if it satisfies the equations*

$$y_i = \sigma_i \varepsilon_i, \tag{1.1}$$

$$\sigma_i^2 = \delta + \alpha y_{i-1}^2 + \beta \sigma_{i-1}^2, \tag{1.2}$$

where δ is a nonnegative function and the integral operators α and β and map nonnegative functions to nonnegative functions, which is in short notation referred to as

$$\delta \geq 0, \quad \alpha \geq 0, \quad \beta \geq 0. \tag{1.3}$$

Further, the innovations

$$(\varepsilon_i: i \in \mathbb{Z}) \text{ are independent and identically distributed processes.} \quad (1.4)$$

Additional assumptions will be introduced were needed. Note that the integral operators α and β are, for $t \in [0, 1]$, defined by $(\alpha x)(t) = \int \alpha(t, s)x(s)ds$ and $(\beta x)(t) = \int \beta(t, s)x(s)ds$, where x is an arbitrary element of $L^2[0, 1]$ and \int is used to mean \int_0^1 . The integral kernel functions $\alpha(s, t)$ and $\beta(s, t)$ are elements of the Hilbert space $L^2[0, 1]^2$ and consequently bounded.

The model given by (1.1) and (1.2) exhibits two time variables. The first is labeled by the integer i and will in the present context often refer to trading day i , even though other time units are possible. The second time variable is labeled by the real-valued t which, without loss of generality, takes values in the unit interval $[0, 1]$. With t , intra-day trading time is parameterized. From this set-up, it should become clear that a modeling of intra-day seasonality may be accommodated by the fGARCH model. Doing this in a functional framework, may lead to efficiency gains over a multivariate modeling approach if the underlying process is sufficiently smooth. Similar to the univariate case argued in Bollerslev (1986), (1.2) should lead to a more flexible modeling of the functional volatility lag structure compared to the fARCH(1) process of Hörmann et al. (2013). The focus is here on the first-order (1,1) case. Even in the univariate case, this is the most widely used GARCH process, used as default in the financial industry because it tends to work well (see, for example, the overview article by Zivot, 2009). Similar arguments should also apply in the more complex functional world.

The paper proceeds as follows. Theoretical properties of the fGARCH(1,1) process defined via (1.1) and (1.2) are discussed in Section 2. This section contains the main contributions of this work, establishing conditions for the existence of the functional stochastic difference equations (1.1) and (1.2). Since one may view the functions y_i and σ_i^2 both as members of the space of square-integrable and continuous functions, two different theorems are stated to cover both cases. The results on the structure of the fGARCH(1,1) process are accompanied by additional results on the estimation of quantities appearing in (1.2). These are given in Section 3. Empirical aspects are highlighted in Section 4. Proofs of the theoretical results are given in Section 5.

2 Structure

This section discusses properties of the fGARCH(1,1) process, in particular notions of unique, strictly stationary solutions to the defining equations (1.1) and (1.2). Two different settings are of interest: solutions in $L^2 = L^2[0, 1]$, the space of square integrable functions on $[0, 1]$, and in $\mathcal{C}[0, 1]$, the space of continuous functions on $[0, 1]$. In both cases, solutions can be characterized through properties of the random integral

kernel

$$\gamma_0(s, t) = \alpha(s, t)\varepsilon_0^2(s) + \beta(t, s).$$

By assumption, $\gamma_0(s, t)$ is an element of $L^2[0, 1]^2$ with probability one and, following Riesz and Sz.–Nagy (1990, p. 148), gives rise to the random integral operator γ_0 defined through $(\gamma x)(t) = \int \gamma_0(s, t)x(s)ds$. Note that $\gamma_0 \geq 0$ (it maps nonnegative functions to nonnegative functions) and denote its (random) Hilbert–Schmidt norm by

$$\|\gamma_0\|_S = \left(\iint \gamma_0^2(s, t)dt ds \right)^{1/2}.$$

In the L^2 case and for the fGARCH(1,1), the Hilbert–Schmidt norm $\|\gamma_0\|_S$ will play the role of the usual top Lyapunov exponent when characterizing the existence of stationary solutions to random difference equations (see, for example, Bouchérol and Picard, 1992). One may also observe the similarity to the univariate GARCH(1,1) case discussed in Nelson (1990).

Theorem 2.1. *Assume that (1.3) and (1.4) hold, that $\delta \in L^2$, and that $\varepsilon_0 \in L^2$ with probability one.*

(i) *If*

$$-\infty \leq \mathbb{E}[\log \|\gamma_0\|_S] < 0, \quad (2.1)$$

then (1.1) and (1.2) have a unique, strictly stationary and nonanticipative solution in L^2 .

(ii) *If there is $\nu > 0$ such that*

$$\mathbb{E}[\|\gamma_0\|_S^\nu] < 1, \quad (2.2)$$

then (1.1) and (1.2) have a unique, strictly stationary and nonanticipative solution in L^2 and $\mathbb{E}[\|\sigma_0^2\|_2^\nu] < \infty$.

The proof of Theorem 2.1 is given in Section 5. If $(y_i : i \in \mathbb{Z})$ is an fARCH(1) process, then $\gamma_0(s, t) = \alpha(s, t)\varepsilon_0^2(s)$. The condition for the existence of a strictly stationary solution for the fARCH(1) equations in L^2 used in Hörmann et al. (2013) can be rewritten as $\mathbb{E}[\|\gamma_0\|_S^\alpha] < 1$. It can then be seen readily from an application of Jensen’s inequality that this condition implies (2.1).

Corollary 2.1. *The following statements are consequences of Theorem 2.1.*

(i) *There is a functional g such that*

$$\sigma_i^2 = g(\varepsilon_{i-1}, \varepsilon_{i-2}, \dots), \quad i \in \mathbb{Z}, \quad (2.3)$$

and therefore $(\sigma_i^2 : i \in \mathbb{Z})$ and $(y_i : i \in \mathbb{Z})$ are Bernoulli shifts.

(ii) *The representation (2.3) implies the ergodicity of $(\sigma_i^2 : i \in \mathbb{Z})$ and $(y_i : i \in \mathbb{Z})$.*

(iii) *Both $(\sigma_i^2 : i \in \mathbb{Z})$ and $(y_i : i \in \mathbb{Z})$ are weakly dependent processes.*

(iv) *For $\ell \in \mathbb{N}$ and $i, j \in \mathbb{Z}$, let $\varepsilon_{i,j}^{(\ell)}$ be independent copies of ε_0 and define*

$$\sigma_{i,\ell}^2 = g(\varepsilon_{i-1}, \dots, \varepsilon_{i-\ell}, \varepsilon_{i-\ell-1, i-\ell}^{(i)}, \varepsilon_{i-\ell-2, i-\ell}^{(i)}, \dots).$$

If (2.2) is satisfied, then there is a constant $0 < \rho < 1$ such that

$$\mathbb{E} [\|\sigma_i^2 - \sigma_{i,\ell}^2\|_2^\nu] = O(\rho^\ell). \quad (2.4)$$

Statement (iv) of the corollary says that, for every ℓ , the sequence $(\sigma_i^2 : i \in \mathbb{Z})$ can be approximated with the ℓ -dependent sequence $(\sigma_{i,\ell}^2 : i \in \mathbb{Z})$ under a geometric rate. Consequently, if (2.2) holds with some $\nu \geq 2$, then the partial sum process $(N^{-1/2} \sum_{i=1}^{\lfloor Nx \rfloor} \{\sigma_i^2 - \mathbb{E}[\sigma_i^2]\})$ satisfies the functional central limit theorem in the space of square integrable processes; see Berkes et al. (2012) and Jirak (2013) for recent theoretical contributions in this area.

If smoother trajectories are required of the fGARCH(1,1) and its volatility process, one may view $(y_i : i \in \mathbb{Z})$ and $(\sigma_i^2 : i \in \mathbb{Z})$ as functional sequences in $\mathcal{C}[0, 1]$. To give the analog of Theorem 2.1 for continuous functions, let $\bar{\gamma}_0 = \int \gamma_0(\cdot, s) ds$ and observe that $\bar{\gamma} \in \mathcal{C}[0, 1]$. The next theorem shows that the role of $\|\gamma_0\|_s$ for square integrable functions is assumed by $\|\bar{\gamma}\|_c = \sup_t |\bar{\gamma}(t)|$ for continuous functions.

Theorem 2.2. *Assume that (1.3) and (1.4) hold, that $\delta \in \mathcal{C}[0, 1]$, and that $\varepsilon_0 \in \mathcal{C}[0, 1]$ with probability one.*

(i) *If*

$$-\infty \leq \mathbb{E} [\log \|\bar{\gamma}_0\|_c] < 0, \quad (2.5)$$

then (1.1) and (1.2) have a unique, strictly stationary and nonanticipative solution in $\mathcal{C}[0, 1]$.

(ii) *If there is $\nu > 0$ such that*

$$\mathbb{E} [\|\bar{\gamma}\|_c^\nu] < 1, \quad (2.6)$$

then (1.1) and (1.2) have a unique, strictly stationary nonanticipative solution in $\mathcal{C}[0, 1]$ and $\mathbb{E} [\|\sigma_0^2\|_c^\nu] < \infty$.

Corollary 2.2. *As in Corollary 2.1, the sequences $(y_i : i \in \mathbb{Z})$ and $(\sigma_i : i \in \mathbb{Z})$ are weakly dependent. Under condition (2.6) there is a constant $0 < \rho < 1$ such that*

$$\mathbb{E} [\|\sigma_i^2 - \sigma_{i,\ell}^2\|_c^\nu] = O(\rho^\ell),$$

where $\sigma_{i,\ell}^2$ is defined in Corollary 2.1.

The proofs of Theorem 2.2 and Corollary 2.2 are given in Section 5. The next section covers an estimation procedure for the fGARCH(1,1) process. Empirical examples are part of Section 4.

3 Estimation

The goal of this section is to obtain an estimate of the volatility equation (1.2), that is, of the function δ and the operators α and β in (1.2). Since the objects involved in the estimation procedure are infinite-dimensional, an M -dimensional class $\Phi_M = \{\varphi_1, \varphi_2, \dots, \varphi_M\}$ of orthonormal functions on $[0, 1]$ is introduced to represent δ ,

α and β in the following way. It is assumed that δ can be represented as a linear combination of the functions in Φ_M , that is,

$$\delta = \sum_{m=1}^M d_m \varphi_m. \quad (3.1)$$

It is further assumed that the integral kernels $\alpha(s, t)$ and $\beta(s, t)$ are elements of the span of $\Phi_M \times \Phi_M$, so that

$$\alpha(t, s) = \sum_{m, m'=1}^M a_{m, m'} \varphi_m(t) \varphi_{m'}(s) \quad \text{and} \quad \beta(t, s) = \sum_{m, m'=1}^M b_{m, m'} \varphi_m(t) \varphi_{m'}(s). \quad (3.2)$$

With these requirements in place, the problem of estimating δ , α and β from a functional sample y_1, y_2, \dots, y_n reduces to estimating the set of real-valued parameters $\{d_m, a_{m, m'}, b_{m, m'} : m, m' = 1, \dots, M\}$. To this end, project first y_1^2, \dots, y_n^2 and $\sigma_1^2, \dots, \sigma_n^2$ onto Φ_M and define the M -dimensional vectors $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,M})^\top$ and $\mathbf{s}_i^2 = (s_{i,1}, \dots, s_{i,M})^\top$ through their entries $y_{i,m} = \langle y_i^2, \varphi_m \rangle$ and $s_{i,m} = \langle \sigma_i^2, \varphi_m \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in L^2 and $^\top$ stands for the transpose of vectors and matrices. Using (1.2) it follows that

$$\mathbf{s}_i^2 = \mathbf{d} + \mathbf{A} \mathbf{y}_{i-1}^2 + \mathbf{B} \mathbf{s}_{i-1}^2, \quad (3.3)$$

where $\mathbf{d} = (d_1, \dots, d_M)^\top$ with $d_m = \langle \delta, \varphi_m \rangle$ and \mathbf{A} and \mathbf{B} are $M \times M$ matrices whose (m, m') th entries are given by $a_{m, m'}$ and $b_{m, m'}$. With this multivariate formulation of the volatility equation, iterating (3.3) yields

$$\mathbf{s}_i^2 = \sum_{\ell=1}^{\infty} \mathbf{B}^{\ell-1} \mathbf{d} + \sum_{\ell=1}^{\infty} \mathbf{B}^{\ell-1} \mathbf{A} \mathbf{y}_{i-\ell}^2, \quad (3.4)$$

where $\mathbf{B}^0 = \mathbf{I}$, the $M \times M$ identity matrix. Note that (3.4) provides a representation for the projections of the unobservable volatilities in terms of the observations. Define the generic equation

$$\tilde{\mathbf{s}}_i^2 = \tilde{\mathbf{s}}_i^2(\underline{\theta}) = \sum_{\ell=1}^{\infty} \underline{\mathbf{B}}^{\ell-1} \underline{\mathbf{d}} + \sum_{\ell=1}^{\infty} \underline{\mathbf{B}}^{\ell-1} \underline{\mathbf{A}} \mathbf{y}_{i-\ell}^2, \quad (3.5)$$

where $\underline{\theta} = (\underline{\mathbf{d}}^\top, \underline{\mathbf{A}}^\top, \underline{\mathbf{B}}^\top)^\top \in \Theta \subset \mathbb{R}^{M \times 2M^2}$. It is clear that $\mathbf{s}_i^2 = \tilde{\mathbf{s}}_i^2(\theta)$, where $\theta = (\mathbf{d}^\top, \mathbf{A}^\top, \mathbf{B}^\top)^\top$ is the true parameter of the fGARCH(1,1) process ($y_i : i \in \mathbb{Z}$). Since this can always be achieved by an appropriate scaling, it is assumed without loss of continuity that

$$\mathbb{E}[\varepsilon_i^2(t)] = 1 \quad \text{for all } t \in [0, 1]. \quad (3.6)$$

If $\mathcal{F}_i = \sigma(\varepsilon_\ell : \ell \leq i)$ denotes the σ -algebra generated by the all innovations up to i , then (3.6) implies that $\mathbb{E}[\mathbf{y}_i^2 | \mathcal{F}_i] = \mathbf{s}_{i-1}^2$ and hence the least square estimator of θ is given as the smallest value of the criterion function $\sum_{i=1}^n \{\mathbf{y}_i^2 - \tilde{\mathbf{s}}_i^2(\underline{\theta})\}^\top \{\mathbf{y}_i^2 - \tilde{\mathbf{s}}_i^2(\underline{\theta})\}$. Since only observe y_1, \dots, y_n are observed, the actual parameter estimator is based on a truncated version of (3.5), namely on

$$\hat{\mathbf{s}}_i^2 = \hat{\mathbf{s}}_i^2(\underline{\theta}) = \sum_{\ell=1}^n \underline{\mathbf{B}}^{\ell-1} \underline{\mathbf{d}} + \sum_{\ell=1}^{i-1} \underline{\mathbf{B}}^{\ell-1} \underline{\mathbf{A}} \mathbf{y}_{i-\ell}^2.$$

This leads to the least squares estimator

$$\hat{\underline{\theta}}_n = \operatorname{argmin} \left\{ S_n(\underline{\theta}) = \sum_{i=2}^n \{ \mathbf{y}_i^2 - \hat{\mathbf{s}}_i^2(\underline{\theta}) \}^\top \{ \mathbf{y}_i^2 - \hat{\mathbf{s}}_i^2(\underline{\theta}) \} : \underline{\theta} \in \Theta \right\},$$

with $\hat{\underline{\theta}}_n = (\hat{\mathbf{d}}_n^\top, \hat{\mathbf{A}}_n^\top, \hat{\mathbf{B}}_n^\top)^\top$. The main result of this section is the strong consistency of $\hat{\underline{\theta}}_n$. Some guarantees to ensure the identifiability of θ are needed and collected next. Let $\| \cdot \|$ denote the Euclidean norm of vectors and matrices.

Assumption 3.1. *It is assumed that*

(A1) y_1^2 is not measurable with respect to \mathcal{F}_0 ;

(A2) \mathbf{A} is nonsingular and $\|\mathbf{B}\| < 1$;

(A3) Θ is a compact set and θ is in the interior of Θ ; and

(A4) there are $0 < c_1$ and $c_2 < 1$ such that $c_1 \leq |\det(\underline{\mathbf{A}})|$ and $\|\underline{\mathbf{B}}\| \leq c_2$ for all $\underline{\theta} = (\underline{\mathbf{d}}^\top, \underline{\mathbf{A}}^\top, \underline{\mathbf{B}}^\top)^\top \in \Theta$.

Part (A1) of Assumption 3.1 means that y_1^2 cannot be predicted almost surely from its past. Parts (A3) and (A4) on the parameter space Θ contain standard regularity conditions.

Theorem 3.1. *Assume that (1.1)–(1.4) and (2.2) hold with $\nu = 1$, that $E[\|\varepsilon_0\|_2^2] < \infty$ and (3.6) is satisfied. Then $\hat{\underline{\theta}}_n$ is strongly consistent for θ under Assumption 3.1, that is,*

$$\hat{\underline{\theta}}_n \rightarrow \theta \quad a.s.$$

as $n \rightarrow \infty$.

The proof of Theorem 3.1 is given in Section 5. Typical choices for the orthonormal class of functions Φ_M include the common Fourier and B -spline bases, but other choices such as wavelets may be entertained as well. If the finite-dimensionality condition (3.2) is not satisfied, then δ , α and β can typically still be well approximated by finite-dimensional functions assuming that M is sufficiently large. While for theoretical considerations M should grow with the sample size, it is worthwhile noting that in practice often a small choice of M (less than, say, 5) will already work reasonably well. This aspect was investigated in the simulations reported in Aue et al. (2015) and is also reported elsewhere in the functional data literature (see Horváth and Kokoszka, 2012).

A different route for utilizing the estimation procedure that leads to a small class Φ_M is to follow the idea of functional principal components analysis (see Horváth and Kokoszka, 2012). For example, it can be proposed to use the eigenfunctions of the covariance kernel

$$D(t, s) = \operatorname{Cov}(y_0^2(t), y_0^2(s)), \quad t, s \in [0, 1],$$

as a basis, so that a few large eigenvalues and their corresponding eigenfunctions capture the most important directions of randomness in y_1^2, \dots, y_n^2 and hence in the unobservable volatilities $\sigma_1^2, \dots, \sigma_n^2$. Since $D(t, s)$ is unknown, the estimation is based on the empirical covariance kernel

$$\hat{D}_n(t, s) = \frac{1}{n} \sum_{i=1}^n \{y_i^2(t) - \bar{y}_n^{(2)}(t)\} \{y_i^2(s) - \bar{y}_n^{(2)}(s)\}, \quad (3.7)$$

where $\bar{y}_n^{(2)} = \frac{1}{n} \sum_{i=1}^n y_i^2$. The spectral decomposition of $\hat{D}(t, s)$ gives rise to the ordered empirical eigenvalues $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_n$ and the corresponding empirical eigenfunctions $\hat{\varphi}_1, \dots, \hat{\varphi}_n$. Therefore, Φ_M may be replaced with the class of estimated eigenfunctions $\hat{\Phi}_M = \{\hat{\varphi}_1, \dots, \hat{\varphi}_M\}$, with appropriately chosen M . The following result shows that consistency can be established also for the functional principal components approach to estimating the fGARCH(1,1) process.

Corollary 3.1. *If $\int \mathbb{E}[y_0^4(t)]dt < \infty$, then the result of Theorem 3.1 is retained if the random class $\hat{\Phi}_M$ is used in place of a deterministic class Φ_M .*

The proof of Corollary 3.1 is given in Section 5.

4 Empirical results

4.1 Simulations

In order to investigate the performance of the least squares procedure in finite samples, functional data was generated from the fGARCH(1,1) process ($y_i: i \in \mathbb{Z}$) whose volatility function (1.2) is specified through the constant intercept function δ given by

$$\delta(t) = .01, \quad t \in [0, 1], \quad (4.1)$$

and the operators α and β defined through their kernel functions

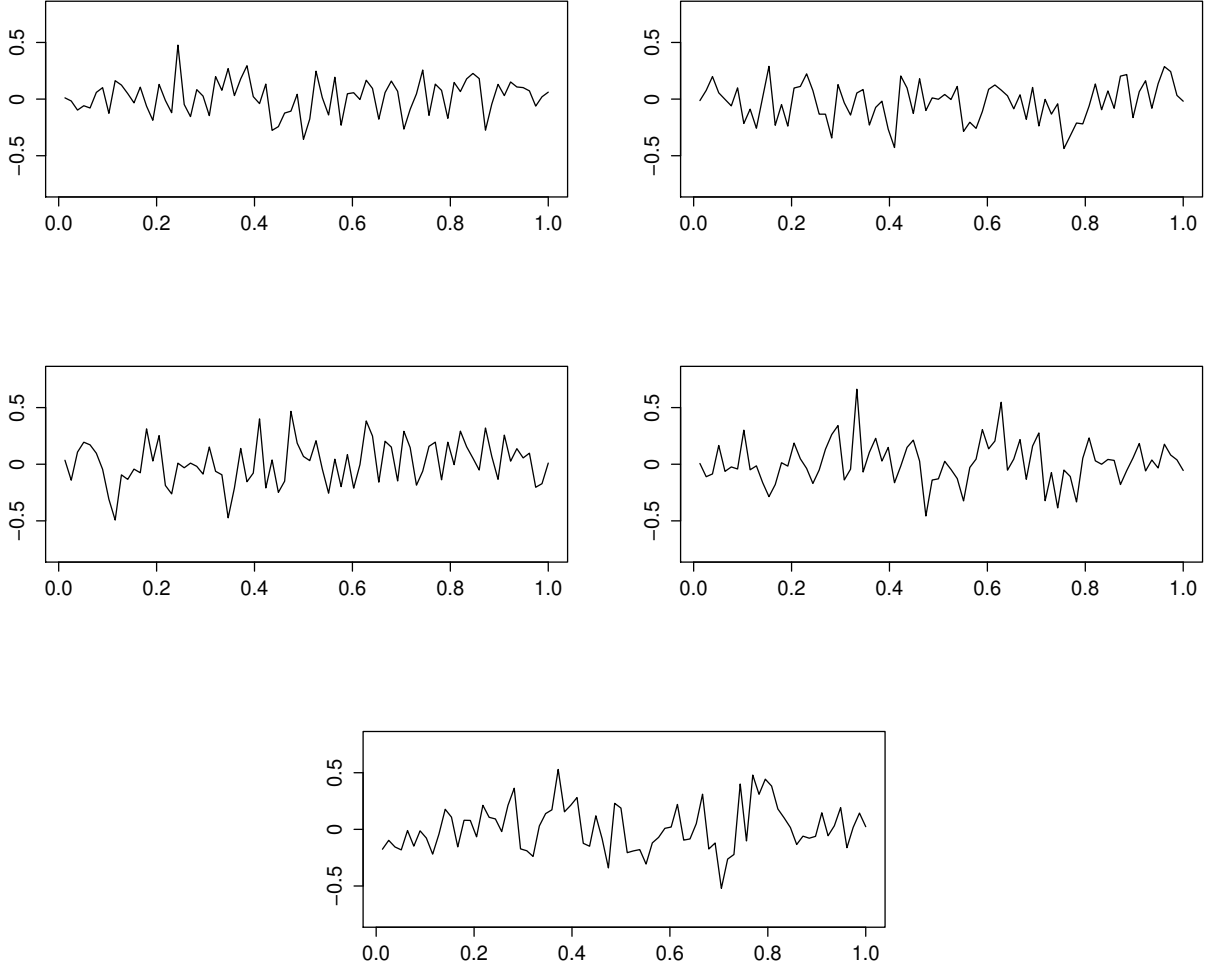
$$\alpha(t, s) = \beta(t, s) = 12t(1-t)s(1-s), \quad t, s \in [0, 1]. \quad (4.2)$$

The innovations ($\varepsilon_i: i \in \mathbb{Z}$) appearing in (1.1) and (1.2) functions are defined as

$$\varepsilon_i(t) = \frac{\sqrt{\log 2}}{2^{200t}} B_i \left(\frac{2^{400t}}{\log 2} \right), \quad t \in [0, 1], \quad (4.3)$$

where ($B_i: i \in \mathbb{Z}$) are independent and identically distributed standard Brownian motions. Lengthy but elementary calculations show that (2.1) holds in case of (4.1)–(4.3) and therefore ($y_i: i \in \mathbb{Z}$) constitutes a strictly stationary process. The recursion in (1.2) was initialized with $\sigma_1^2 = \delta$ and the first 1000 simulated curves were discarded as burn-in values since to get close to the stationary solution. Figure 4.1 displays the first five simulated functions after the burn-in.

Figure 4.1: Five consecutive fGARCH(1,1) observations (read row-wise starting from the top-left) generated from model (4.1)–(4.3).



Random samples of size $n = 300, 600$ and 1200 were obtained from the data generating process in equations (4.1)–(4.3). Note that in this setting $\alpha(t, s) = \beta(t, s) = .4\varphi_1(t)\varphi_1(s)$ with the normalized function φ_1 given by

$$\varphi_1(t) = \sqrt{30}t(1 - t), \quad t \in [0, 1].$$

Equation (3.2) shows then that the operator-related parameters to estimate are $a_{11} = b_{11} = .4$. Note that the intercept function $\delta \equiv 0.01$ is not in the span of φ_1 and therefore does not satisfy (3.1), so that the corresponding least squares estimate becomes a proxy for $d_1 = \langle \delta, \varphi_1 \rangle = \sqrt{30}/600 \approx .009$. The estimation procedure was run 1000 times with $M = 1$ yielding the outcomes reported in Table 4.1.

An alternative method for estimating δ can be obtained as follows. Observing that if $\hat{\alpha}$ and $\hat{\beta}$ are asymp-

totically consistent estimators for α and β in the L^2 -sense, then

$$\tilde{\delta}_n = \bar{y}_n^{(2)} - (\hat{\alpha} + \hat{\beta})\bar{y}_n^{(2)}$$

is an asymptotically consistent estimator for δ , where $\bar{y}_n^{(2)} = \frac{1}{n} \sum_{i=1}^n y_i^2$. In the present simulation setting, there was little difference between $\hat{\delta} = \hat{d}_1\varphi$ and $\tilde{\delta}$, so only results for the first choice are presented here.

Table 4.1: Estimates \hat{d}_1 , \hat{a}_{11} and \hat{b}_{11} with $M = 1$ and $\varphi_1(t) = \sqrt{30t(1-t)}$ in model (4.1)–(4.3) for different sample sizes n . The row labeled by ∞ shows the population values.

n	\hat{d}_1	\hat{a}_{11}	\hat{b}_{11}
300	0.013	0.420	0.306
600	0.011	0.412	0.344
1200	0.010	0.408	0.369
∞	0.009	0.400	0.400

To illustrate the estimation method using estimated functional principal components, whose large-sample behavior is given in Corollary 3.1, $n = 1000$ functional observations were generated from the fGARCH(1,1) model (4.1)–(4.3). The largest eigenvalue of $\hat{D}_{1,000}$, the empirical covariance kernel defined in (3.7), explains about 70% of the variability in the squared functions y_i^2 , while the second-largest eigenvalue only accounts for 4%. First, the estimation procedure was applied using only the first empirical eigenfunction $\hat{\varphi}_1$. The resulting estimated kernels $\hat{\alpha}_{1000}(t, s)$ and $\hat{\beta}_{1000}(t, s)$ are shown in Figure 4.2 along with the population kernels $\alpha(t, s) = \beta(t, s)$ given in (4.3). The experiment was then repeated using two or more eigenfunction but no qualitative improvement in the shape of the estimators $\alpha_{1000}(t, s)$ and $\beta_{1000}(t, s)$ was found. As before, \hat{d}_{1000} estimates the sum of the first coefficients of the expansion of δ in the empirical eigenbasis $\hat{\varphi}_1, \hat{\varphi}_2, \dots$

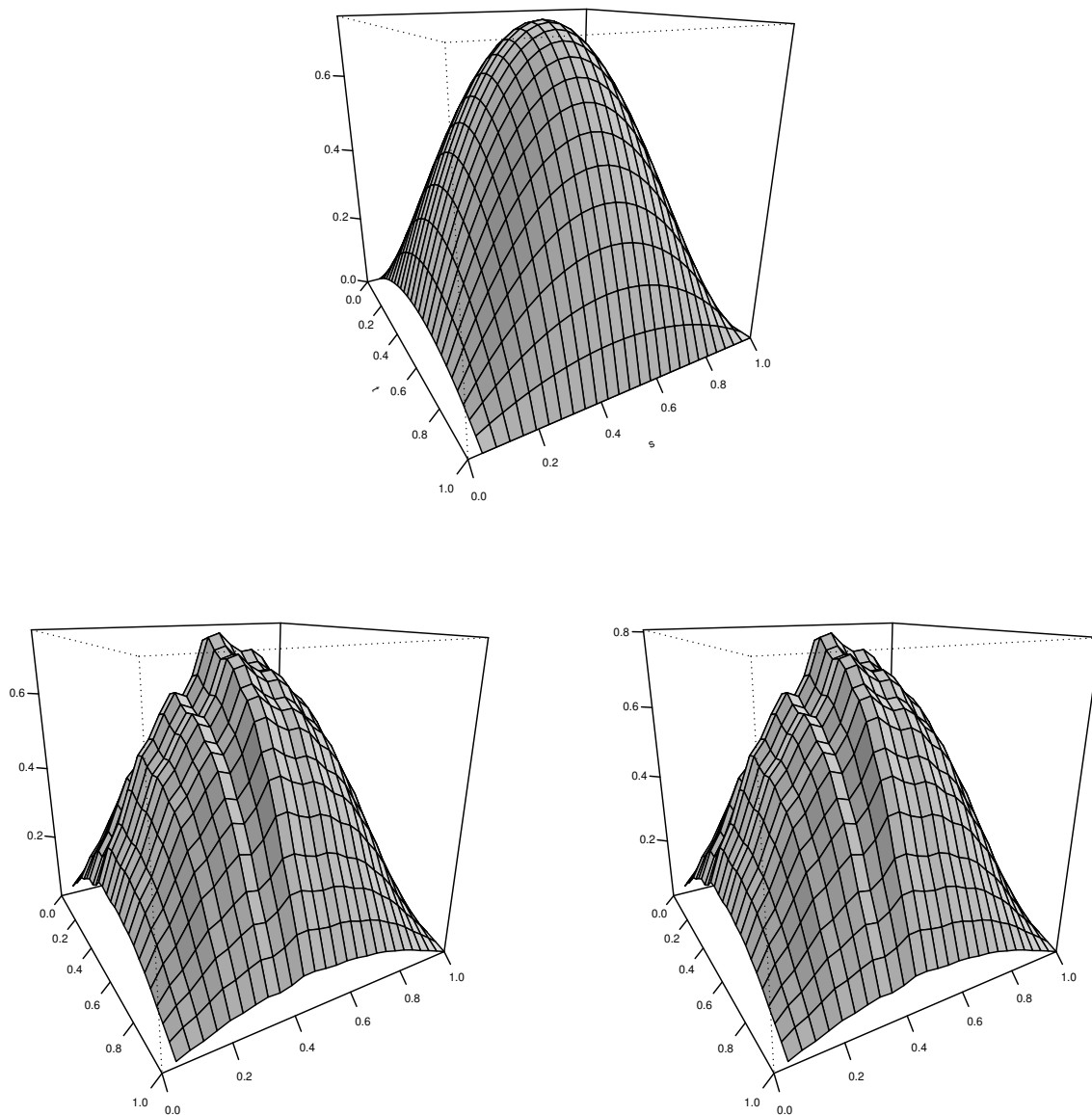
4.2 Application

The data example considered in this subsection is on intra-day log-returns of an exchange-traded fund (ETF), namely the popular SPDR S&P 500 EDF (SPY) designed to track the S&P 500 index. Let then $p_i(t)$ denote the value of the SPY ETF on day i at time t . Following Cyree et al. (2004) one can set up the intra-day log-returns

$$y_i(t) = \log p_i(t) - \log p_i(t - h),$$

where h is a time window typically corresponding to 1, 5, or 15 minutes. In the following, only $h = 5$ and therefore 5-minute log-returns are considered. This is done for convenience because the price of the SPY ETF is recorded in five-minute resolution, for example, at the website <https://quantquote.com>, and the 5-minute log-return process (y_i) becomes easily computable. Since the SPY trades from 9:30am to 4:00pm, there are 78 measurements per day. The resulting high-dimensionality of the intra-day observations make the

Figure 4.2: Plots of $\alpha(t, s) = \beta(t, s)$ of (4.2) (top panel) and corresponding estimates $\hat{\alpha}_{1000}(t, s)$ and $\hat{\beta}_{1000}(t, s)$ (lower panel) based on the data generating process (4.1)–(4.3).



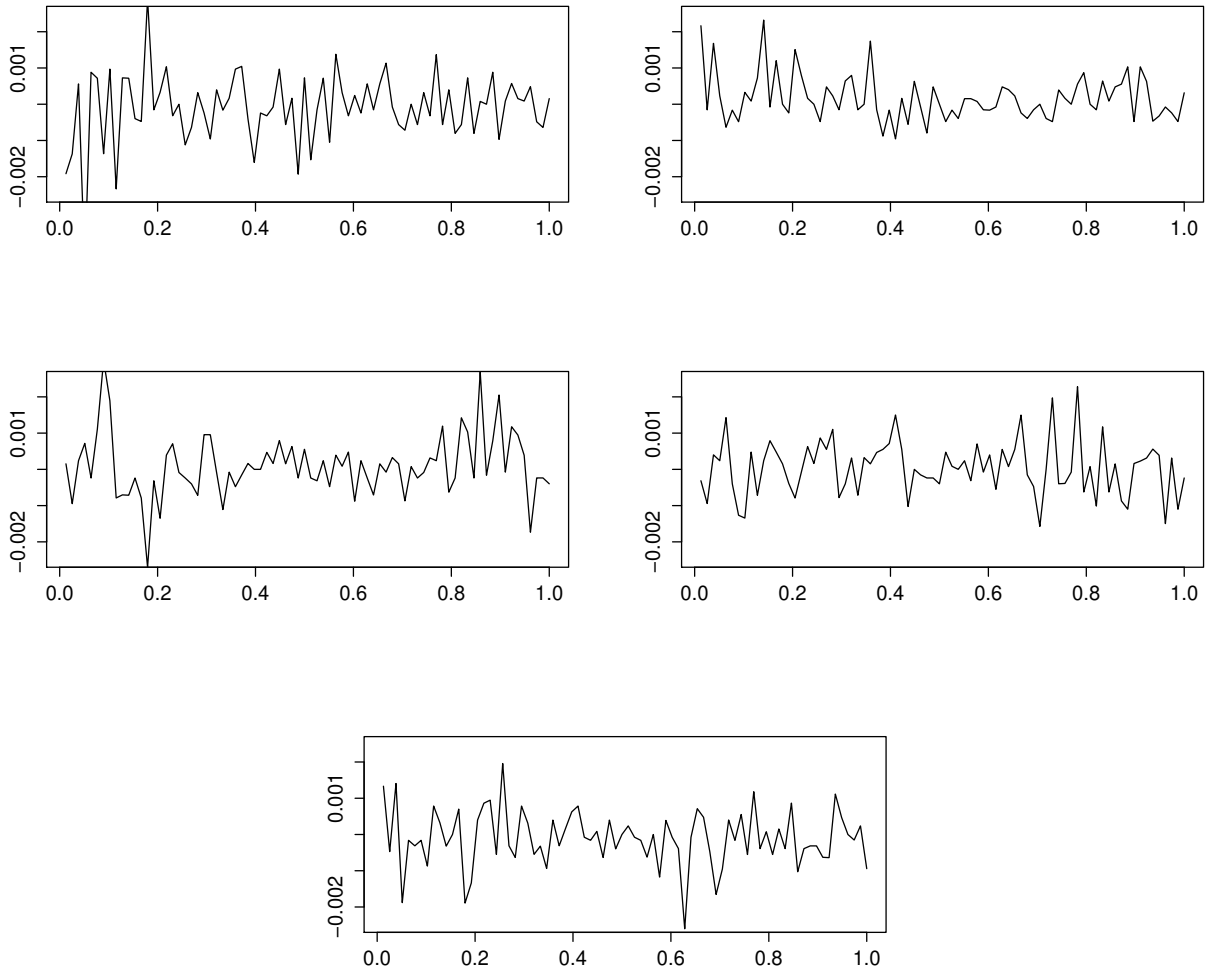
application of any multivariate ARCH or GARCH model impractical and a functional alternative is suggested in the following. Note that the volatility of the log-returns on day i at time t is then represented by

$$\sigma_i^2(t) = \text{Var}(y_i(t)|\mathcal{F}_{i-1}),$$

where \mathcal{F}_{i-1} denotes the the available information until the end of day $i - 1$. Figure 4.3 shows the graphs of the 5-minute log-returns of the SPY data for five consecutive trading days in July 2006.

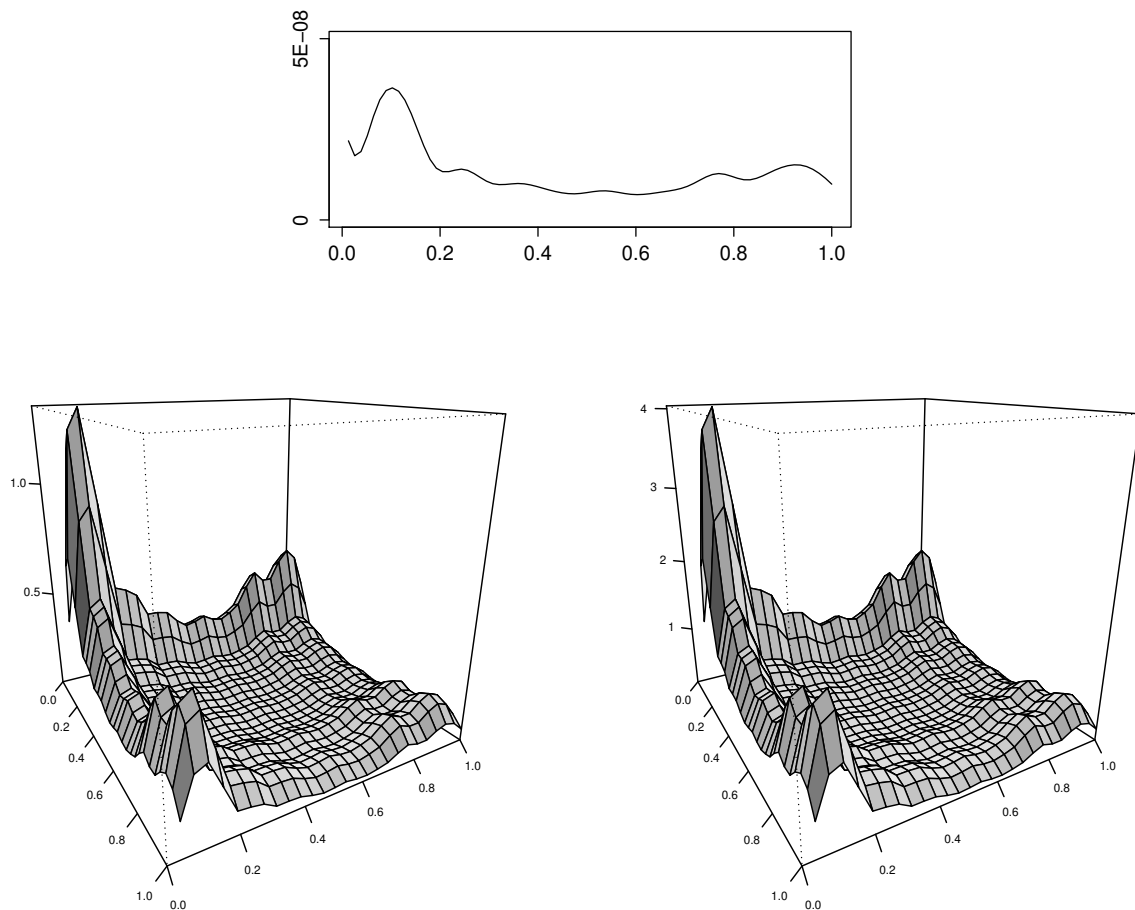
The full data set considered here covers the trading days between January 2, 2003 and December 30,

Figure 4.3: Plots of 5-minute log-returns of the SPY ETF from July 21 and 24–27, 2006 (read row-wise starting from the top-left).



2006. SPY is not traded on weekends and bank holidays. There were eight half trading days during this four year span and due to possibly different market behavior these eight curves were excluded from the analysis, resulting in the fitting of an fGARCH(1,1) process to $n = 999$ functional observations. The least squares estimation procedure was carried out using Φ_M consisting of the empirical eigenfunctions associated with the M largest eigenvalues of the covariance operator \hat{D}_{999} . Figure 4.4 shows the estimates for the intercept function δ and integral kernels corresponding to the operators α and β for the case $M = 1$. As in the simulated data example, increasing the number of eigenfunctions used, did not qualitatively improve the estimators. It should be noted that $\hat{\delta}$ has a global maximum shortly after trading opens in the morning and a further local maximum at the end of the day, corroborating similar findings in other papers.

Figure 4.4: Estimates for intercept function δ (top), and kernels $\alpha(t, s)$ (bottom left) and $\beta(t, s)$ (bottom right) based on the SPY ETF data.



5 Proofs

5.1 Proofs of Section 2 results

Using (1.1), the volatility equation (1.2) can, for any $t \in [0, 1]$, be written as

$$\sigma_i^2(t) = \delta(t) + \int \gamma_{i-1}(t, s) \sigma_{i-1}^2(s) ds, \quad (5.1)$$

with integral kernel

$$\gamma_{i-1}(t, s) = \alpha(t, s) \varepsilon_{i-1}^2(s) + \beta(t, s).$$

Iterating (5.1) backward yields

$$\begin{aligned} \sigma_i^2(t) &= \delta(t) + \int \gamma_{i-1}(t, s_1) \left[\delta(s_1) + \int \gamma_{i-2}(s_1, s_2) \sigma_{i-2}^2(s_2) ds_2 \right] ds_1 \\ &= \delta(t) + \int \gamma_{i-1}(t, s_1) \delta(s_1) ds_1 + \iint \gamma_{i-1}(t, s_1) \gamma_{i-2}(s_1, s_2) \sigma_{i-2}^2(s_2) ds_2 ds_1 \end{aligned}$$

$$\begin{aligned}
&= \delta(t) + \int \gamma_{i-1}(t, s_1) \delta(s_1) ds_1 + \iint \gamma_{i-1}(t, s_1) \gamma_{i-2}(s_1, s_2) \delta(s_2) ds_2 ds_1 \\
&\quad + \iiint \gamma_{i-1}(t, s_1) \gamma_{i-2}(s_1, s_2) \gamma_{i-3}(s_2, s_3) \sigma_{i-3}^2(s_3) ds_3 ds_2 ds_1,
\end{aligned}$$

and so on. For $i \in \mathbb{Z}$, let $\Gamma_{i,0}$ denote the identity operator and, for $i \in \mathbb{Z}$ and $k \in \mathbb{N}$, let the random integral operators $\Gamma_{i,k}$ be defined by the equation

$$(\Gamma_{i,k} x)(t) = \int \cdots \int \gamma_{i-1}(t, s_1) \gamma_{i-2}(s_1, s_2) \cdots \gamma_{i-k}(s_{k-1}, s_k) x(s_k) ds_k \cdots ds_2 ds_1.$$

Then, it follows that, for any $m \in \mathbb{N}$,

$$\sigma_i^2 = \sum_{k=0}^{m-1} \Gamma_{i,k} \delta + \Gamma_{i,m} \sigma_{i-m}^2. \quad (5.2)$$

It can be seen from the definition of $\Gamma_{i,k}$ that

$$\Gamma_{i,k} = \Gamma_{i,1} \circ \Gamma_{i-1,1} \circ \cdots \circ \Gamma_{i-k+1,1}, \quad (5.3)$$

where \circ denotes the composition of operators. The foregoing identifies

$$\sigma_i^2 = \sum_{k=0}^{\infty} \Gamma_{i,k} \delta, \quad (5.4)$$

as candidate for the nonanticipative solution of (1.1) and (1.2). It remains to be shown that the sum on the right-hand side of equation (5.4) is well defined with probability one. This will be verified in the following.

Proof of Theorem 2.1. (i) Consider the convergence of the sum in (5.4) in L^2 , the space of square integrable functions on $[0, 1]$. Since

$$(\Gamma_{i,1} x)(t) = \int \gamma_{i-1}(t, s) x(s) ds$$

defines an integral operator in L^2 , it follows from Riesz and Sz.–Nagy (1990, p. 148) that

$$\|\Gamma_{i,1}\|_{\mathcal{N}} \leq \|\gamma_{i-1}\|_2, \quad (5.5)$$

where $\|\cdot\|_{\mathcal{N}}$ denotes the operator norm on L^2 . Using representation (5.3) and inequality (5.5) leads to

$$\|\Gamma_{i,k}\|_{\mathcal{N}} \leq \prod_{\ell=1}^k \|\gamma_{i-\ell}\|_2. \quad (5.6)$$

(The usual arrangement $\prod_{i \in \emptyset} = 1$ applies here and in the following.) Since $\|\gamma_\ell\|_2$ are independent and identically distributed random variables, the strong law of large numbers implies that

$$\frac{1}{k} \sum_{\ell=1}^k \log \|\gamma_{i-\ell}\|_2 \rightarrow \mathbb{E}[\log \|\gamma_0\|_2] \quad \text{a.s.} \quad (5.7)$$

Hence, in view of (2.1),

$$P \left\{ \sum_{k=0}^{\infty} (\Gamma_{i,k} \delta)(t) \in L^2[0, 1] \right\} = 1.$$

It is easy to see that that the infinite sum on the right side of (5.4) satisfies (5.1) and hence (1.1) and (1.2).

To prove the uniqueness of the solution to (1.1) and (1.2) assume that $\bar{\sigma}_i^2$ is another such solution. Using the backward recursions as in (5.2) shows that, for all $m \in \mathbb{N}$,

$$\bar{\sigma}_i^2 = \sum_{k=0}^{m-1} \Gamma_{i,k} \delta + \Gamma_{i,m} \bar{\sigma}_{i-m}^2. \quad (5.8)$$

Hence, for all $m \in \mathbb{N}$,

$$\left\| \bar{\sigma}_i^2 - \sum_{k=0}^{\infty} \Gamma_{i,k} \delta \right\|_2 \leq \left\| \sum_{k=m}^{\infty} \Gamma_{i,k} \delta \right\|_2 + \|\Gamma_{i,m} \bar{\sigma}_{i-m}^2\|_2. \quad (5.9)$$

Assumption (2.1) implies for the first norm on the right-hand side of the latter equation that, as $m \rightarrow \infty$,

$$\left\| \sum_{k=m}^{\infty} \Gamma_{i,k} \delta \right\|_2 \leq \sum_{k=m}^{\infty} \|\Gamma_{i,k}\|_{\mathcal{N}} \|\delta\|_2 \rightarrow 0 \quad \text{a.s.}$$

Similarly,

$$\|\Gamma_{i,m} \bar{\sigma}_{i-m}^2\|_2 \leq \|\Gamma_{i,m}\|_{\mathcal{N}} \|\bar{\sigma}_{i-m}^2\|_2 \xrightarrow{P} 0,$$

which follows since $\|\Gamma_{i,m}\|_{\mathcal{N}} \rightarrow 0$ a.s. and, due to the weak stationarity of $(\bar{\sigma}_i: i \in \mathbb{Z})$, the sequence $(\|\bar{\sigma}_{i-m}^2\|_2: i \in \mathbb{Z})$ is bounded in probability. Therefore

$$P \left\{ \left\| \bar{\sigma}_i - \sum_{k=0}^{\infty} \Gamma_{i,k} \delta \right\|_2 = 0 \right\} = 1,$$

which proves the first part of the theorem.

(ii) Note that

$$\|\sigma_0^2\|_2 \leq \sum_{k=0}^{\infty} \|\Gamma_{i,k} \delta\|_2 \leq \|\delta\|_2 \sum_{k=0}^{\infty} \|\Gamma_{i,k}\|_{\mathcal{N}} \leq \|\delta\|_2 \sum_{k=0}^{\infty} \prod_{\ell=1}^k \|\gamma_{i-\ell}\|_2.$$

Consequently Minkowski's inequality (cf. Hardy et al., 1959, pp. 24–26) implies that, for $\nu \geq 1$,

$$(\mathbb{E}[\|\sigma_0^2\|_2^\nu])^{1/\nu} \leq \|\delta\|_2 \sum_{k=0}^{\infty} \left(\mathbb{E} \left(\prod_{\ell=1}^k \|\gamma_{i-\ell}\|_2 \right)^\nu \right)^{1/\nu} \leq \|\delta\|_2 \sum_{k=0}^{\infty} (\mathbb{E}[\|\gamma_0\|_2^\nu])^{k/\nu}$$

and, for $0 < \nu \leq 1$,

$$\mathbb{E}[\|\sigma_0^2\|_2^\nu] \leq \|\delta\|_2^\nu \sum_{k=0}^{\infty} (\mathbb{E}[\|\gamma_0\|_2^\nu])^k$$

The second part of the theorem is proved. \square

Proof of Corollary 2.1. (i) The Bernoulli shift representation in (2.3) is an immediate consequence of (5.4).

(ii) Ergodicity is implied by the work of Stout (1974).

(iv) For $j \in \mathbb{N}$, let

$$\gamma_{i,\ell,i-\ell-j}(t, s) = \alpha(t, s) \varepsilon_{i,\ell,i-\ell-j}^2(s) + \beta(t, s).$$

and, for $k \geq \ell$, let the operators $\Gamma_{i,\ell,k}$ be defined through

$$\begin{aligned} (\Gamma_{i,\ell,k} x)(t) &= \int \cdots \int \gamma_{i-1}(t, s_1) \gamma_{i-2}(s_1, s_2) \cdots \gamma_{i-\ell-1}(s_{\ell-2}, s_{\ell-1}) \\ &\quad \times \gamma_{i,\ell,i-\ell}(s_{\ell-1}, s_\ell) \gamma_{i,\ell,i-k}(s_{k-1}, s_k) x(s_k) ds_k ds_{k-1} \cdots ds_1. \end{aligned}$$

Setting

$$\sigma_{i,\ell}^2 = \sum_{k=0}^{\ell-1} \Gamma_{i,k} \delta + \sum_{k=\ell}^{\infty} \Gamma_{i,\ell,k} \delta,$$

it can be seen that

$$\mathbb{E} \left[\|\sigma_i^2 - \sigma_{i,\ell}^2\|_2^\nu \right] \leq 2^\nu \mathbb{E} \left[\left\| \sum_{k=\ell}^{\infty} \Gamma_{i,k} \delta \right\|_2^\nu \right] + 2^\nu \mathbb{E} \left[\left\| \sum_{k=\ell}^{\infty} \Gamma_{i,\ell,k} \delta \right\|_2^\nu \right].$$

It follows from the proof of Theorem 2.1 that

$$E \left[\left\| \sum_{k=\ell}^{\infty} \Gamma_{i,k} \delta \right\|_2^\nu \right] \leq \begin{cases} \|\delta\|_2^\nu \left(\sum_{k=\ell}^{\infty} \{\mathbb{E}[\|\gamma_0\|_2^\nu]\}^{k/\nu} \right)^\nu, & \text{if } \nu \geq 1, \\ \|\delta\|_2^\nu \sum_{k=\ell}^{\infty} \{\mathbb{E}[\|\gamma_0\|_2^\nu]\}^k, & \text{if } 0 < \nu \leq 1. \end{cases}$$

Observing that

$$\mathbb{E} \left[\left\| \sum_{k=\ell}^{\infty} \Gamma_{i,k} \delta \right\|_2^\nu \right] = \mathbb{E} \left[\left\| \sum_{k=\ell}^{\infty} \Gamma_{i,\ell,k} \delta \right\|_2^\nu \right]$$

and that $E[\|\gamma_0\|_2^\nu] < 1$ by assumption, the proof of (2.4) is complete. \square

Proof of Theorem 2.2. (i) Recall relations (5.2)–(5.4). It follows from the definition of $\Gamma_{i,k}$ that

$$\|\Gamma_{i,k} \delta\|_e \leq \|\delta\|_e \prod_{\ell=1}^k \|\tilde{\gamma}_{i-\ell}\|_e$$

with $\tilde{\gamma}_j(t) = \int \gamma_j(t, s) ds$. Hence

$$\left\| \sum_{k=0}^{\infty} \Gamma_{i,k} \delta \right\|_e \leq \|\delta\|_e \sum_{k=0}^{\infty} \prod_{\ell=1}^k \|\tilde{\gamma}_{i-\ell}\|_e.$$

The strong law of large numbers yields that

$$\frac{1}{k} \sum_{\ell=1}^k \log \|\tilde{\gamma}_{i-\ell}\|_e \rightarrow \mathbb{E}[\log \|\tilde{\gamma}_0\|_e] \quad \text{a.s.,}$$

implying immediately that, under assumption (2.5)

$$P\left\{\sum_{k=0}^{\infty}\Gamma_{i,k}\delta\in\mathcal{C}[0,1]\right\}=1.$$

Thus the existence of a solution to (1.1) and (1.2) is proven. Assume that $(\bar{\sigma}_i^2: i\in\mathbb{Z})$ is an other solution to (1.1) and (1.2) in $\mathcal{C}[0,1]$. Since (5.8) holds, it follows as in (5.9) that, for all $m\in\mathbb{N}$,

$$\left\|\bar{\sigma}_i-\sum_{k=0}^{\infty}\Gamma_{i,k}\delta\right\|_{\mathcal{C}}\leq\left\|\sum_{k=m}^{\infty}\Gamma_{i,k}\delta\right\|_{\mathcal{C}}+\|\Gamma_{i,m}\bar{\sigma}_{i-m}^2\|_{\mathcal{C}}.$$

If (2.5) is satisfied, then, as $m\rightarrow\infty$,

$$\left\|\sum_{k=m}^{\infty}\Gamma_{i,k}\right\|_{\mathcal{C}}\leq\|\delta\|_{\mathcal{C}}\sum_{k=m}^{\infty}\|\Gamma_{i,k}\|_{\mathcal{C}}\rightarrow 0\quad\text{a.s.},$$

using that $(1/m)\log\|\Gamma_{i,m}\|_{\mathcal{C}}\rightarrow\mathbb{E}[\log\|\bar{\gamma}_0\|_{\mathcal{C}}]<0$ with probability one. By the assumed weak stationarity of the function series $(\bar{\sigma}_i: i\in\mathbb{Z})$, the univariate sequence $(\|\bar{\sigma}_{i-m}^2\|_{\mathcal{C}}: i\in\mathbb{Z})$ is bounded in probability and therefore

$$\|\Gamma_{i,m}\bar{\sigma}_{i-m}^2\|_{\mathcal{C}}\leq\|\Gamma_{i,m}\|_{\mathcal{C}}\|\bar{\sigma}_{i-m}^2\|_{\mathcal{C}}\xrightarrow{P}0.$$

Consequently,

$$P\left\{\left\|\bar{\sigma}_i-\sum_{k=0}^{\infty}\Gamma_{i,k}\delta\right\|_{\mathcal{C}}=0\right\}=1,$$

and the uniqueness of the solution to (1.1) and (1.2) is established.

(ii) The proof is similar to the proof of part (ii) of Theorem 2.1. □

Proof of Corollary 2.2. The proof follows along the line of arguments used to establish Remark 2.1 and the details are thus omitted. □

5.2 Proofs of Section 3 results

Proof of Theorem 3.1. First observe that, by (1.1), $\langle y_i^2, \varphi_m \rangle = \langle \sigma_i^2 \varepsilon_i^2, \varphi_m \rangle = \langle \sigma_i^2, \varphi_m \rangle + \langle \sigma_i^2 (\varepsilon_i^2 - 1), \varphi_m \rangle$.

Due to the independence of ε_i and σ_i , $y_{i-\ell}$, $\ell\in\mathbb{N}$, (3.6) and the stationarity of $(y_i: i\in\mathbb{Z})$ imply that

$$\mathbb{E}[(\mathbf{y}_i^2 - \tilde{\mathbf{s}}_i^2)^\top (\mathbf{y}_i^2 - \tilde{\mathbf{s}}_i^2)] = \sum_{m=1}^M E\langle \sigma_0^2 (\varepsilon_0^2 - 1), \varphi_m \rangle^2 + M(\underline{\theta}), \quad (5.10)$$

where

$$M(\underline{\theta}) = \mathbb{E}\{[\mathbf{s}_0^2 - \tilde{\mathbf{s}}_0^2(\underline{\theta})]^\top [\mathbf{s}_0^2 - \tilde{\mathbf{s}}_0^2(\underline{\theta})]\}, \quad \underline{\theta} \in \Theta.$$

Since $M(\underline{\theta}) \geq 0$, $M(\underline{\theta}) = 0$ and the first term on the right-hand side of (5.10) is constant as a function of $\underline{\theta}$, it suffices to show that

$$M(\underline{\theta}) > 0 \quad \text{if } \underline{\theta} \neq \theta. \quad (5.11)$$

Assume now that (5.11) does not hold, so that there is $\theta_* \neq \theta$, $\theta_* \in \Theta$ such that $M(\theta_*) = 0$. But this means that $\mathbf{s}_0^2 = \tilde{\mathbf{s}}_0^2(\theta_*)$ with probability one. Using (3.4) and (3.5), it can be concluded that

$$\sum_{\ell=1}^{\infty} \mathbf{B}^{\ell-1} \mathbf{d} + \sum_{\ell=1}^{\infty} \mathbf{B}^{\ell-1} \mathbf{A} \mathbf{y}_{-\ell}^2 = \sum_{\ell=1}^{\infty} \mathbf{B}_*^{\ell-1} \mathbf{d}_* + \sum_{\ell=1}^{\infty} \mathbf{B}_*^{\ell-1} \mathbf{A}_* \mathbf{y}_{-\ell}^2, \quad (5.12)$$

with $\theta_* = (\mathbf{d}_*, \mathbf{A}_*, \mathbf{B}_*)$. If ℓ^* is the smallest integer such that $\mathbf{B}^{\ell^*-1} \mathbf{A} \neq \mathbf{B}_*^{\ell^*-1} \mathbf{A}_*$, then

$$(\mathbf{B}^{\ell^*-1} \mathbf{A} - \mathbf{B}_*^{\ell^*-1} \mathbf{A}_*) \mathbf{y}_{-\ell^*}^2 = \sum_{\ell=1}^{\infty} \mathbf{B}^{\ell-1} \mathbf{d} - \sum_{\ell=1}^{\infty} \mathbf{B}_*^{\ell-1} \mathbf{d}_* + \sum_{\ell=\ell^*+1}^{\infty} (\mathbf{B}^{\ell-1} \mathbf{A} - \mathbf{B}_*^{\ell-1} \mathbf{A}_*) \mathbf{y}_{-\ell}^2$$

and therefore $\mathbf{y}_{-\ell^*}^2$ is measurable with respect to $\mathcal{F}_{-\ell^*-1}$ which contradicts part (A1) of Assumption 3.1 due to stationarity. Hence $\mathbf{B}^\ell \mathbf{A} = \mathbf{B}_*^\ell \mathbf{A}_*$ for all $\ell \geq 0$ resulting in $\mathbf{A} = \mathbf{A}_*$ and consequently $\sum_{\ell=1}^{\infty} \mathbf{B}^{\ell-1} \mathbf{A} = \sum_{\ell=1}^{\infty} \mathbf{B}_*^{\ell-1} \mathbf{A}$, in turn implying that $(\mathbf{I} - \mathbf{B})^{-1} \mathbf{A} = (\mathbf{I} - \mathbf{B}_*)^{-1} \mathbf{A}$. Thus $\mathbf{B} = \mathbf{B}_*$. Since $\mathbf{A} = \mathbf{A}_*$ and $\mathbf{B} = \mathbf{B}_*$, (5.12) shows that $\sum_{\ell=1}^{\infty} \mathbf{B}^{\ell-1} \mathbf{d} = \sum_{\ell=1}^{\infty} \mathbf{B}_*^{\ell-1} \mathbf{d}_*$ and hence $(\mathbf{I} - \mathbf{B})^{-1} \mathbf{d} = (\mathbf{I} - \mathbf{B}_*)^{-1} \mathbf{d}_*$, completing the proof of $\mathbf{d} = \mathbf{d}_*$. By contradiction, (5.11) is established.

The ergodic theorem (cf. Breiman, 1968) yields that, for all $\underline{\theta} \in \Theta$,

$$\frac{1}{n} \sum_{i=2}^n (\mathbf{y}_i^2 - \tilde{\mathbf{s}}_{i-1}^2(\underline{\theta}))^\top (\mathbf{y}_i^2 - \tilde{\mathbf{s}}_{i-1}^2(\underline{\theta})) \rightarrow \sum_{m=1}^M \mathbb{E} [\langle \sigma_0^2(\varepsilon_0^2 - 1), \varphi_m \rangle^2 + M(\underline{\theta})] \quad \text{a.s.}$$

as $n \rightarrow \infty$. Observing that the derivatives of $\tilde{\mathbf{s}}_0^2$ have a uniformly bounded expected value on $\underline{\theta} \in \Theta$, standard arguments (see, for example, the proof of the uniform law of large numbers in Ferguson, 1996) show that

$$\sup_{\underline{\theta} \in \Theta} \left| \frac{1}{n} \sum_{i=2}^n \{ \mathbf{y}_i^2 - \tilde{\mathbf{s}}_{i-1}^2(\underline{\theta}) \}^\top \{ \mathbf{y}_i^2 - \tilde{\mathbf{s}}_{i-1}^2(\underline{\theta}) \} - \left(\sum_{m=1}^M \mathbb{E} \langle \sigma_0^2(\varepsilon_0^2 - 1), \varphi_m \rangle^2 + M(\underline{\theta}) \right) \right| \rightarrow 0 \quad \text{a.s.}$$

Using (3.5) and the assumption that $\|\underline{\mathbf{B}}\| \leq \rho < 1$ on Θ gives

$$\sup_{\underline{\theta} \in \Theta} \left| \frac{1}{n} \sum_{i=2}^n \{ \mathbf{y}_i^2 - \tilde{\mathbf{s}}_{i-1}^2(\underline{\theta}) \}^\top \{ \mathbf{y}_i^2 - \tilde{\mathbf{s}}_{i-1}^2(\underline{\theta}) \} - \frac{1}{n} \sum_{i=2}^n \{ \mathbf{y}_i^2 - \hat{\mathbf{s}}_{i-1}^2(\underline{\theta}) \}^\top \{ \mathbf{y}_i^2 - \hat{\mathbf{s}}_{i-1}^2(\underline{\theta}) \} \right| \rightarrow 0 \quad \text{a.s.}$$

Thus,

$$\sup_{\underline{\theta} \in \Theta} \left| \frac{1}{n} \sum_{i=2}^n \{ \mathbf{y}_i^2 - \hat{\mathbf{s}}_{i-1}^2(\underline{\theta}) \}^\top \{ \mathbf{y}_i^2 - \hat{\mathbf{s}}_{i-1}^2(\underline{\theta}) \} - \left(\sum_{m=1}^M \mathbb{E} \langle \sigma_0^2(\varepsilon_0^2 - 1), \varphi_m \rangle^2 + M(\underline{\theta}) \right) \right| \rightarrow 0 \quad \text{a.s.} \quad (5.13)$$

The results in (5.11) and (5.13) imply the consistency in Theorem 3.1 via standard arguments. \square

Proof of Corollary 3.1. The approximability established in Corollary 2.1 implies that $\|\hat{D}_N - D\| \rightarrow 0$ in probability and therefore

$$\max_{1 \leq m \leq M} \|\hat{\varphi}_m - \hat{\zeta}_m \varphi_m\| \xrightarrow{P} 0, \quad (n \rightarrow \infty),$$

where the $\hat{\zeta}_m$'s are random signs, assuming that $\lambda_1 > \dots > \lambda_M > \lambda_{M+1}$ are the eigenvalues of D in decreasing order (see Horváth and Kokoszka, 2012). Hence the consistency of Theorem 3.1 remains true when the functions in $\hat{\Phi}_M = \{\hat{\varphi}_1, \dots, \hat{\varphi}_M\}$ are used to set up (3.2). \square

References

- [1] Aue, A., Dubart Nourinho, D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association* **110**, 378–392.
- [2] Aue, A., Hörmann, S., Horváth, L., and Hušková, M. (2014). Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica* **24**, 1043–1073.
- [3] Bauwens, I., Laurent, S. and Rombouts, J.V.K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* **21**(2006), 79–109.
- [4] Berkes, I., Horváth, L. and Rice, G. (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Processes and Their Applications* **123**, 385–403.
- [5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.
- [6] Bosq, D. (2010). *Linear Processes in Function Spaces*. Springer-Verlag, New York.
- [7] Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* **52**, 115–127.
- [8] Breiman, L. (1968). *Probability*. Addison Wesley.
- [9] Cyree, K.K., Griffiths, M.D. and Winters, D.B. (2004). An empirical examination of intraday volatility in euro-dollar rates. *The Quarterly Review of Economics and Finance* **44**, 44–57.
- [10] Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.
- [11] Engle, R.F. (2002). Dynamic conditional correlation: a simple class of multivariate GARCH models. *Journal of Business of Economic Statistics* **20**, 339–350.
- [12] Engle, R.F. and Kroner, K. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**, 122–150.
- [13] Engle, R.F. and Mezrich, J. (1996). GARCH for groups. *Risk* **9**, 36–40.
- [14] Ferguson, T. (1996). *A Course in Large Sample Theory*. Chapman and Hall.
- [15] Francq, C. and Zakoian, J.-M. (2010). *GARCH Models*. Wiley, Chichester, U.K.

- [16] Francq, C., Horváth, L. and Zakoian, J.-M. (2011). Merits and drawbacks of variance Targeting in GARCH models. *Journal of Financial Econometrics* **9**, 619–656.
- [17] Francq, C., Horváth, L. and Zakoian, J.-M. (2015+). Variance targeting estimation of multivariate GARCH models. *Journal of Financial Econometrics*, in press.
- [18] Hardy, G.H., Littlewood, J.E. and Pólya, G. (1959). *Inequalities (2nd ed.)*. Cambridge University Press, Cambridge, U.K.
- [19] Hörmann, S., Horváth, L. and Reeder, R. (2013). A functional version of the ARCH model. *Econometric Theory* **29**, 267–288.
- [20] Horváth, L. and Kokoszka, P. (2012), *Inference for Functional Data with Applications*. Springer-Verlag, New York.
- [21] Jirak, M. (2013). On weak invariance principles for sums of dependent random functionals. *Statistics & Probability Letters* **83**, 2291–2296.
- [22] Nelson, D.B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**, 318–334.
- [23] Ramsay, J.O. and Silvermann, B.W. (2005). *Functional Data Analysis (2nd ed.)*. Springer-Verlag, New York.
- [24] Riesz, F. and Sz.–Nagy, B. (1990). *Functional Analysis*. Dover, New York.
- [25] Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models. In: *Handbook of Financial Time Series (Eds: T.G. Anderson et al.)*, pp. 201–229. Springer-Verlag, New York.
- [26] Stout, W.F. (1974). *Almost Sure Convergence*. Academic Press, New York.
- [27] Zivot, E. (2009). Practical issues in the analysis of univariate GARCH models. In: *Handbook of Financial Time Series (Eds: T.G. Anderson et al.)*, pp. 113–155. Springer-Verlag, New York.