Journal of Econometrics 209 (2019) 353–375

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

# Functional GARCH models: The quasi-likelihood approach and its applications

Clément Cerovecki [a], Christian Francq [b], Siegfried Hörmann [c], Jean-Michel Zakoïan [d,*]

[a] Université libre de Bruxelles, Boulevard du Triomphe, CP210 B-1050 Brussels - Belgium Office: NO O.9.211, Belgium
[b] CREST and University of Lille, BP 60149, 59653 Villeneuve d'Ascq cedex, France
[c] Graz University of Technology, Kopernikusgasse 24/III, Austria
[d] University of Lille and CREST, 5 Avenue Henry Le Chatelier, 91120 Palaiseau, France

ARTICLE INFO

*Article history:*
Received 17 January 2018
Received in revised form 14 September 2018
Accepted 7 January 2019
Available online 30 January 2019

*JEL classification:*
C13
C32
C58

*Keywords:*
Functional time series
High-frequency volatility models
Intraday returns
Functional QMLE
Stationarity of functional GARCH

ABSTRACT

The increasing availability of high frequency data has initiated many new research areas in statistics. Functional data analysis (FDA) is one such innovative approach towards modelling time series data. In FDA, densely observed data are transformed into curves and then each (random) curve is considered as one data object. A natural, but still relatively unexplored, context for FDA methods is related to financial data, where high-frequency trading currently takes a significant proportion of trading volumes. Recently, articles on functional versions of the famous ARCH and GARCH models have appeared. Due to their technical complexity, existing estimators of the underlying functional parameters are moment based—an approach which is known to be relatively inefficient in this context. In this paper, we promote an alternative quasi-likelihood approach, for which we derive consistency and asymptotic normality results. We support the relevance of our approach by simulations and illustrate its use by forecasting realised volatility of the S&P100 Index.


## 1. Introduction

Financial time series modelling is of great importance in monitoring the evolution of prices, stock indexes or exchange rates and to predict future developments, such as the risk associated to certain asset allocations. Risk is very much related to the volatility of the financial process and, hence, models for volatility are of special importance. A milestone in volatility modelling has been set by Engle (1982), with the introduction of the now-famous and widely-used ARCH model. Many extensions followed this groundbreaking work, most notably the GARCH model by Bollerslev (1986) which allows a more parsimonious fit in comparison to ARCH processes. The success of these models is founded on their mathematical feasibility and on their ability to feature many of the *stylised facts* that researchers have been observing in empirical investigations of financial data. In particular, the models are able to capture a non-constant conditional variance of time series. For details on GARCH models, see for example, Francq and Zakoïan (2011) and the references therein.

In practical applications, GARCH models and their variations are adequate for daily or weekly return data. But, due to the availability of high-frequency financial time series and their importance for the financial industry, it is desirable to provide

* Corresponding author.
*E-mail addresses:* clement.cerovecki@ulb.ac.be (C. Cerovecki), christian.francq@univ-lille3.fr (C. Francq), shoermann@tugraz.at (S. Hörmann), zakoian@ensae.fr (J.-M. Zakoïan).

https://doi.org/10.1016/j.jeconom.2019.01.006
0304-4076/

corresponding models and adequate statistical methodology for data that are given at a higher resolution. In this paper, we adopt the theory of *functional time series* to approach this challenge. A functional time series is a sequence of observations $(X_t : 1 \le t \le n)$, where each random object $X_t$ is a curve $(X_t(u) : u \in [0, 1])$. The interval $[0, 1]$ is chosen for convenience and does not impose any restriction on generality. In our context it represents intraday time. For example, $X_t(u)$ might denote the price of an asset on day $t$ at intraday time $u$. If we then consider the log-returns $y_t(u) = \log X_t(u) - \log X_t(u - \tau)$ on some $\tau$-interval or the intraday log-increments $\tilde{y}_t(u) = \log X_t(u) - \log X_t(0)$, then it seems plausible that such common transformations yield stationary functional processes (as processes in the discrete time $t$), in which case a variety of tools can be employed for inference on the intraday pattern.

Functional time series methods have received increasing attention during the past few years. To give a small sample of some very recent articles with many further references we refer to the following papers: Hörmann and Kokoszka (2010) and Eichler and van Delft (2017) for structural results, Horváth et al. (2014) and Aue and Klepsch (2017) for inferential procedures, Paparoditis (2017) and Zhu and Politis (2017) for functional time series bootstrapping methods and Aue et al. (2015) and Klepsch and Klüppelberg (2017) for forecasting algorithms.

In this paper, we consider some adequate functional models to describe, for instance, the functional time series $(y_t)$ or $(\tilde{y}_t)$ as defined above. The first attempt to generalise GARCH models to *functional time series* was made in Hörmann et al. (2013), where a functional version of the ARCH(1) was proposed. Later, this model was extended in Aue et al. (2017) to a functional GARCH(1,1). Both models rely on recurrence equations with unknown operators and curves. As for the estimation of these quantities, Hörmann et al. (2013) proposed a moment-based estimator and showed its consistency. Their approach allows to deal with a fully functional (and potentially infinite-dimensional) parameter space. The situation is more complicated in the GARCH context. Aue et al. (2017) proposed a least squares estimator based on the recursive empirical volatility. This approach comes at a price: the authors have to reduce the functional model to a multivariate model via some dimension reduction to a fixed finite dimension. Moreover, it is known from scalar GARCH theory that the least-squares estimators lack efficiency, see e.g. Theorem 6.4 and refer to Tables 6.2 and 7.1 for numerical comparison in Francq and Zakoïan (2011).

In this paper, *we propose an estimator inspired by the classical GARCH QML (Quasi-Maximum Likelihood) method* (Section 3). The definition of a QMLE is far from straightforward in that context, because a likelihood cannot be written for curves. Our estimator is based on the projection of the squared process onto a set of non-negative valued baseline functions. We give regularity conditions for *consistency and asymptotic normality*. As a side result, we obtain the consistency and asymptotic normality of a semi-strong (i.e. with non-iid innovations) multivariate Constant Conditional Correlation (CCC-GARCH, see the Appendix for more details). We also obtain a sufficient condition for existence of stationary functional GARCH processes (Section 2.2) which generalises (Aue et al., 2017). We use the top Lyapunov exponent formulation, and our condition is very similar to the sufficient and necessary condition that can be obtained in the finite dimensional case. Our results also extend to higher order models, i.e. functional GARCH($p$, $q$). In terms of application, we use our model to predict *realised volatility*, which is an important volatility measure (Section 5.2).

The rest of the paper is organised as follows: in Section 2, we introduce the model equations, some notations and discuss the stationarity. In Section 3, we introduce our estimation procedure and detail its asymptotic properties. In Section 4, we extend our consistency results to infinite-dimensional models. The subsequent sections deal with practical aspects of the implementation and some empirical illustrations which demonstrate the superiority of the QMLE compared to existing methods. Technical results and proofs are given in the Appendix.

## 2. Functional GARCH($p$, $q$) model

### 2.1. Preliminaries

For convenience we first review notation. We denote by $H$ the Hilbert space of square integrable functions with domain $[0, 1]$. It will serve as the basic space on which the functional observation, that is considered in this paper, is defined. The Hilbert space is equipped with inner product $\langle \cdot, \cdot \rangle$ and the resulting norm $\| \cdot \|$. If $x$ and $y$ are both functions of $H$ (respectively, vectors of $\mathbb{R}^d$), then we denote by $xy$ their point-wise (resp. component-wise) product. We denote by $\mathcal{L}(H)$ the space of bounded linear operators on $H$ and use bold notation for its elements. Hence, for $\boldsymbol{\alpha} \in \mathcal{L}(H)$, $x \in H$ and $u \in [0, 1]$ we have that $\boldsymbol{\alpha}(x)$ is the image in $H$ of $\boldsymbol{\alpha}$ applied to $x$, whereas $x(u)$ is the real-valued image of the function $x$ evaluated at $u$. Moreover, we use the standard convention for combining operators, i.e. that $\boldsymbol{\alpha}\boldsymbol{\beta} := \boldsymbol{\alpha} \circ \boldsymbol{\beta}$ and $\boldsymbol{\alpha}^2 := \boldsymbol{\alpha} \circ \boldsymbol{\alpha}$ for $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{L}(H)$. We recall that $\mathcal{L}(H)$, equipped with the usual operator norm $\|\boldsymbol{\alpha}\| := \sup_{\|x\| \le 1} \|\boldsymbol{\alpha}(x)\|$, is a Banach space. This norm is sub-multiplicative, i.e. $\|\boldsymbol{\alpha}\boldsymbol{\beta}\| \le \|\boldsymbol{\alpha}\| \, \|\boldsymbol{\beta}\|$. In some places we also make use of the supremum norm: $\|x\|_\infty = \inf\{a > 0, \ |x(u)| < a, \text{ for } \lambda\text{-almost every } u \in [0, 1]\}$. For $x, y \in H$, we define the operator $x \otimes y := x\langle \cdot, y \rangle$.

We define the subspaces $H^+ = \{x \in H : x(u) \ge 0, \text{ for almost every } u \in [0, 1]\}$ and $H^+_* = \{x \in H : x(u) > 0, \text{ for almost every } u \in [0, 1]\}$. Let $\mathcal{K}(H)$ denote the space of kernel operators on $H$, i.e. if $\boldsymbol{\alpha} \in \mathcal{K}(H)$ then there exists a function $K_{\boldsymbol{\alpha}} : [0, 1] \times [0, 1] \to \mathbb{R}$ such that $\boldsymbol{\alpha}(x)(u) = \int K_{\boldsymbol{\alpha}}(u, v)x(v)dv$. For simplicity, we will often write $\int$ instead of $\int_0^1$. Let $\mathcal{L}^+(H)$ denote the space of operators which map $H^+$ onto $H^+$ and note that an operator $\boldsymbol{\alpha} \in \mathcal{K}^+(H) := \mathcal{L}^+(H) \cap \mathcal{K}(H)$, if and only if its kernel $K_{\boldsymbol{\alpha}}(\cdot, \cdot)$ is non-negative.

For any integer $k \ge 2$, the product space $H^k = H \times \cdots \times H$ naturally inherits the Hilbert space structure by defining its scalar product as $\langle x, y \rangle = \sum_{i=1}^k \langle x_i, y_i \rangle$, for $x, y \in H^k$. In this context, it will be useful to represent elements and operators as $k$-dimensional vectors with values in $H$ and $k \times k$ matrices with values in $\mathcal{L}(H)$, respectively. For example, if $k = 2$, we consider the operator

$$\boldsymbol{\alpha} : x \longmapsto \begin{pmatrix} \boldsymbol{\alpha}_{11} & \boldsymbol{\alpha}_{12} \\ \boldsymbol{\alpha}_{21} & \boldsymbol{\alpha}_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} := \begin{pmatrix} \boldsymbol{\alpha}_{11}(x_1) + \boldsymbol{\alpha}_{12}(x_2) \\ \boldsymbol{\alpha}_{21}(x_1) + \boldsymbol{\alpha}_{22}(x_2) \end{pmatrix}$$

where $x = (x_1, x_2)^\top \in H^2$ and $\boldsymbol{\alpha}_{11}, \boldsymbol{\alpha}_{12}, \boldsymbol{\alpha}_{21}$ and $\boldsymbol{\alpha}_{22} \in \mathcal{L}(H)$.

We are now ready to introduce our general model.

**Definition 1.** Let $(\eta_t)_{t\in\mathbb{Z}}$ be a sequence of i.i.d. random elements of $H$. A functional GARCH($p, q$) process $(y_t)_{t\in\mathbb{Z}}$ is defined as a stationary solution of the equations

$$y_t = \sigma_t \eta_t, \tag{2.1}$$

$$\sigma_t^2 = \delta + \sum_{i=1}^{q} \boldsymbol{\alpha}_i\,(y_{t-i}^2) + \sum_{j=1}^{p} \boldsymbol{\beta}_j\,(\sigma_{t-j}^2), \tag{2.2}$$

where $\delta \in H_*^+$ and $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p \in \mathcal{K}^+(H)$. Such a solution is called non-anticipative if $\sigma_t = \sigma(\eta_{t-1}, \eta_{t-2}, \ldots)$ for some measurable function $\sigma$.

Under the assumption that $E(\eta_t(u)) = 0$ and $E(\eta_t^2(u)) = 1$, the variable $\sigma_t^2(u)$ can be interpreted as the volatility at day $t$ and intraday time $u$, i.e. the variance of the return $y_t(u)$ conditional upon the sigma algebra $\mathcal{F}_{t-1}$ generated by $(\eta_s)_{s\le t-1}$. Note that this volatility may depend on all past returns, not only on those corresponding to intraday time $u$ of the previous days. For instance, let $p = 0, q = 1$ (ARCH(1)), and suppose that $\boldsymbol{\alpha}$ is a kernel operator, with constant kernel $K_{\boldsymbol{\alpha}}(u, v) := a$, then the pattern of the intraday volatility $\sigma_t^2(u) = \delta(u) + a\int y_{t-1}^2(v)dv$ is essentially given by that of $\delta$. Moreover, it depends on the previous day through the so-called 'integrated volatility' given by the integral. If now, the kernel has the form $K_{\boldsymbol{\alpha}}(u, v) = a\phi(u - v)$, where $\phi$ denotes a density function with mode at zero, we have $\sigma_t^2(u) = \delta(u) + a\int \phi(u - v)y_{t-1}^2(v)dv$ and the volatility of intraday-time $u$ is mainly driven by the volatility of the previous day 'around' time $u$. It is, thus, clear that Model (2.1) allows for a great flexibility through the choice of the operators $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and the pattern of the intercept $\delta$.

A key feature of the GARCH model is that it captures well the dynamics of volatility observed in financial data. In our *functional* setting, we propose the following interpretation of the volatility curves. For any fixed $u \in [0, 1]$, we have that

$$P\left(|y_t(u)| < c \mid \mathcal{F}_{t-1}\right) = P\left(|\eta_t(u)| < c/\sigma_t(u) \mid \mathcal{F}_{t-1}\right) = 1 - \alpha, \tag{2.3}$$

if we take $c = \sigma_t(u) \cdot Q_{1-\alpha/2}^{\eta(u)}$. Consider, for example, a process with Gaussian innovations $(\eta_t)$ such that $\mathrm{Var}(\eta_t(u)) = 1$ for all $u \in [0, 1]$. We can then interpret the region $\{[-2\sigma_t(u), 2\sigma_t(u)] : u \in [0, 1]\}$ as the prediction interval of $y_t(u)$ at (approximate) level $\alpha = .05$. We show these curves and their estimation in Fig. 1 (in a setting that will be described below) at two different scales: 7 and 100 days, respectively. The data generating process is described in Section 5.1.2. On the first figure we observe the sensitivity to shocks of the volatility curves. On the second figure we can observe the persistence of the volatility curves on a larger scale.

One interest of the functional GARCH model is that it allows for prediction of the next day's volatility curve. At the end of day $t-1$, the whole volatility curve of day $t$ can be predicted. It is, thus, possible to predict the realised volatility $\sum_{j=1}^{\lfloor 1/\tau \rfloor} y_t^2(j\tau)$ for some given time unit $\tau \in (0, 1)$, or any other realised measure of volatility. This will be illustrated in Section 5.2.

### 2.2. Existence of stationary solutions

In light of Definition 1, an evident question concerns the existence of a strictly stationary and non-anticipative solution to the functional GARCH equations. To respond to this problem, we first observe that our model equations can be conveniently summarised in the following state–space form:

$$\underline{z}_t = \underline{b}_t + \boldsymbol{\Psi}_t(\underline{z}_{t-1}), \tag{2.4}$$

where $\underline{b}_t, \underline{z}_t \in H^{p+q}$ and $\boldsymbol{\Psi}_t \in \mathcal{L}(H^{p+q})$ are defined as

$$\underline{b}_t = \left(\eta_t^2\delta, 0, \ldots, 0, \delta, 0, \ldots, 0\right)', \qquad \underline{z}_t = \left(y_t^2, \ldots, y_{t-q+1}^2, \sigma_t^2, \ldots, \sigma_{t-p+1}^2\right)',$$

and

$$\boldsymbol{\Psi}_t = \begin{pmatrix} \boldsymbol{\Upsilon}_t\boldsymbol{\alpha}_1 & \ldots & \boldsymbol{\Upsilon}_t\boldsymbol{\alpha}_{q-1} & \boldsymbol{\Upsilon}_t\boldsymbol{\alpha}_q & \boldsymbol{\Upsilon}_t\boldsymbol{\beta}_1 & \ldots & \boldsymbol{\Upsilon}_t\boldsymbol{\beta}_{p-1} & \boldsymbol{\Upsilon}_t\boldsymbol{\beta}_p \\ \boldsymbol{I}_H & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ & \ddots & & & & & & \\ \mathbf{0} & & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cdots & \boldsymbol{I}_H & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\alpha}_1 & \ldots & \boldsymbol{\alpha}_{q-1} & \boldsymbol{\alpha}_q & \boldsymbol{\beta}_1 & \ldots & \boldsymbol{\beta}_{p-1} & \boldsymbol{\beta}_p \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \boldsymbol{I}_H & \cdots & \mathbf{0} & \mathbf{0} \\ & & & & & \ddots & & \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{I}_H & \mathbf{0} \end{pmatrix}.$$

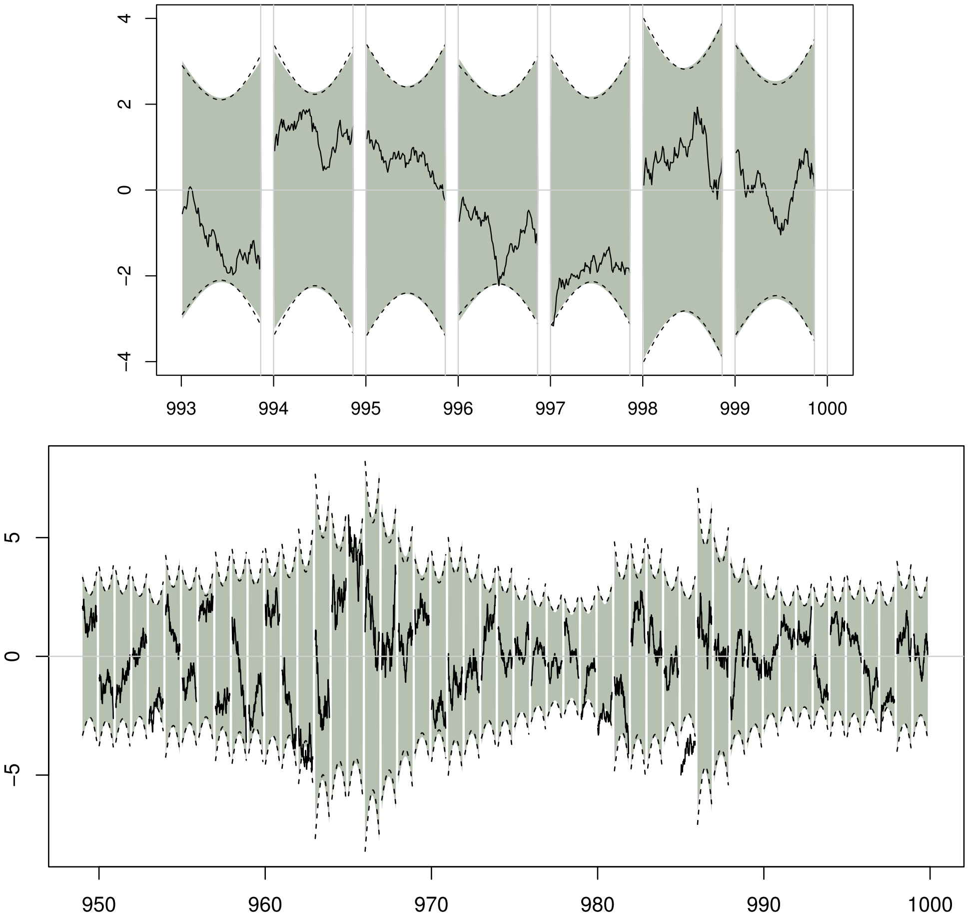


**Fig. 1.** Solid lines represent the simulated process $y_t$, the shaded area is the region $\{[-2\sigma_t(u), 2\sigma_t(u)]: u \in [0,1]\}$. The dashed lines are estimators $\pm 2\widetilde{\sigma}_t(\hat{\theta})(u)$.

Here, the operator $\boldsymbol{\Upsilon}_t$ is the pointwise multiplication by $\eta_t^2$, i.e. $H \ni x \longmapsto x\eta_t^2$. All $\mathbf{0}$'s in the definition of the matrix $\boldsymbol{\Psi}_t$ are meant to be zero-operators.

Now, we introduce a mild technical assumption which we impose for the rest of the paper:

$$E \log^+ \|\eta_0^2\|_\infty < \infty, \tag{2.5}$$

where $\log^+(u) = \log(\max(1,u))$. By this assumption, it follows that $\|\eta_t^2\|_\infty < \infty$ a.s. and, hence, the linear operator $\boldsymbol{\Upsilon}_t$ is almost surely bounded. Indeed, we have that

$$\|\boldsymbol{\Upsilon}_t(x)\| = \|x\eta_t^2\| = \left(\int x^2(u)\eta_t^4(u)du\right)^{1/2} \leq \|\eta_t^2\|_\infty \|x\|, \quad \text{for any } x \in H,$$

thus

$$\|\boldsymbol{\Upsilon}_t\| \leq \|\eta_t^2\|_\infty. \tag{2.6}$$

For the sake of a light notation, we will now also use $\|\cdot\|$ for the norm on $H^{p+q}$ as well as for the operator norm of $\mathcal{L}(H^{p+q})$. Its respective meaning will be clear from the context. From assumption (2.5) it is easily deduced that $E\log^+\|\boldsymbol{\Psi}_1\| < \infty$. Moreover, the sequence $(\boldsymbol{\Psi}_t)$ is i.i.d. and our norm on $\mathcal{L}(H^{p+q})$ is sub-multiplicative. Hence, according to Theorem 6 in Kingman (1973) we have that

$$\gamma := \lim_{t\to\infty} \frac{1}{t} E(\log\|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_1\|) = \inf_{t\geq 1}\frac{1}{t}E(\log\|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_1\|) \tag{2.7}$$

$$= \lim_{t\to\infty}\frac{1}{t}\log\|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_1\|, \ a.s. \tag{2.8}$$

The coefficient $\gamma \in [-\infty, +\infty)$ is called the *top Lyapunov exponent* of the sequence $(\boldsymbol{\Psi}_t)_{t\in\mathbb{Z}}$.

**Theorem 1.** *Under (2.5), a sufficient condition for the existence of a unique strictly stationary and non-anticipative solution to (2.1)–(2.2) is* $\gamma < 0$.

**Remark 1.** It would be interesting to see if the condition $\gamma < 0$ is also necessary for the existence of a strictly stationary solution to (2.1)–(2.2). The situation in the functional context is more complicated when compared to multivariate analysis. In the multivariate setup, one would argue that for some appropriately chosen matrix norm $\|\cdot\|_*$ we have that $\|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_{t-k+1}(\underline{b}_{t-k})\|_* \to 0$ (which is, of course, necessary for the convergence of (A.6)) will imply $\|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_{t-k+1}\|_* \to 0$. In a second step, one uses contraction properties of random matrices in order to conclude. In the infinite-dimensional setup, however, norms are not equivalent, and choosing a different norm will also give a different value for the exponent $\gamma$. Extending such results to linear operators is beyond the scope of this paper.

In the next corollary, we specialise to the case of the functional GARCH(1,1) process in order to obtain a slightly more explicit result.

**Corollary 1.** *When $p = q = 1$, a sufficient condition for existence of a strictly stationary and non-anticipative solution to (2.1)–(2.2) is that*

$$E\log\left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{t-1}+\boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{1}+\boldsymbol{\beta})\right\| < 0, \quad \textit{for some } t \geq 1.$$

In their recent paper, Aue et al. (2017) obtained the condition

$$E\log\|\boldsymbol{\alpha}\boldsymbol{\Upsilon}_0+\boldsymbol{\beta}\|_{\mathcal{S}} < 0 \tag{2.9}$$

to guarantee a strictly stationary solution of functional GARCH(1,1) equations. Here, $\|\boldsymbol{\gamma}_0\|_{\mathcal{S}}$ is the Hilbert–Schmidt norm. Note that the Hilbert–Schmidt norm is dominating the operator norm and, hence,

$$\gamma \leq \frac{1}{t}E\log\left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{t}+\boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{1}+\boldsymbol{\beta})\right\| \leq E\log\left\|\boldsymbol{\alpha}\boldsymbol{\Upsilon}_0+\boldsymbol{\beta}\right\| \leq E\log\|\boldsymbol{\alpha}\boldsymbol{\Upsilon}_0+\boldsymbol{\beta}\|_{\mathcal{S}}.$$

In general these inequalities are strict, which shows that our condition is milder than that of Aue et al. (2017).

In the next proposition, we provide a sufficient condition for $E[y_t^2(u)] < \infty$ and equivalently $E[\sigma_t^2(u)] < \infty$, for all $u \in [0, 1]$. We denote by $\rho(\boldsymbol{A})$ the spectral radius of the operator $\boldsymbol{A}$.

**Proposition 1.** *Under (2.5) and if $\gamma < 0$, then a sufficient condition for the existence of a pointwise second-order stationary solution to (2.1)–(2.2) is that $\rho(E\boldsymbol{\Psi}_0) < 1$.*

We conclude this section with a result, which will be useful for statistical inference, but it also has its own interest.

**Proposition 2.** *Assume that $E\|\eta_0^2\|_\infty^\tau < \infty$ for some $\tau \in (0,1)$, $\gamma < 0$ and that $(y_t)_{t\in\mathbb{Z}}$ is a stationary solution to (2.1)–(2.2). Then there exists $s \in (0,\tau)$ such that $E\|y_t^2\|^s < \infty$ and $E\|\sigma_t^2\|^s < \infty$.*

## 3. Estimation

A difficulty in estimating FDA models is that the concept of likelihood cannot be easily generalised to the infinite dimensional setup. For this reason, QML estimation is not really defined in this framework. We propose an estimator which, though it cannot be related to any likelihood for the aforementioned reason, is directly inspired from the QMLE in the standard GARCH model. The functional QMLE will be shown to be consistent in a semi-parametric framework in which the distribution of $\eta_t$ does not need to be specified.

### *3.1. Parametrisation*

From observations $(y_t)_{1\leq t\leq n}$ of curves satisfying Model (2.1)–(2.2), we consider inference on the parameters $\delta, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q$ and $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p$. In order to guarantee identifiability of the model, we impose

$$E[\eta_0^2(u)] = 1, \ \forall u \in [0, 1]. \tag{3.1}$$

An example of a stationary Gaussian process $(\eta_0(u))_{u\in[0,1]}$ satisfying (3.1) is the Ornstein–Uhlenbeck process given by $\eta_0(u) = e^{-u/2}W_0(e^u)$, where $W_0(\cdot)$ is the standard Brownian motion. This process has autocovariance function $\mathrm{Cov}(\eta_0(u+v), \eta_0(v)) = e^{-u/2}$. In general, however, we do not require either Gaussianity or "intraday-stationarity" of $\eta_t$.

We begin by assuming a specific parametrisation for Model (2.1)–(2.2). Let $\varphi_1, \ldots, \varphi_M$ be linearly independent baseline functions in $H^+$. We assume that there exists a vector $d = (d_1, \ldots, d_M)'$ in $\mathbb{R}^M$, and matrices $A_i = (a_{k,\ell}^{(i)})$ and $B_j = (b_{k,\ell}^{(j)})$ in $\mathbb{R}^{M\times M}$ such that

$$\delta = \sum_{k=1}^{M} d_k\varphi_k, \quad \boldsymbol{\alpha}_i = \sum_{k,\ell=1}^{M} a_{k,\ell}^{(i)}\,\varphi_k\otimes\varphi_\ell \quad \text{and} \quad \boldsymbol{\beta}_j = \sum_{k,\ell=1}^{M} b_{k,\ell}^{(j)}\,\varphi_k\otimes\varphi_\ell, \tag{3.2}$$

for $i = 1, \dots q$, and $j = 1, \dots p$. The coefficients are chosen such that $\delta(u)$ is positive and that $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ belong to $\mathcal{K}_H^+$.[1] Define the parameter

$$\theta = \mathrm{vec}\big(d, A_1, \dots, A_q, B_1, \dots, B_p\big) \in \mathbb{R}^{M+(p+q)M^2}. \tag{3.3}$$

The model (2.1)–(2.2) is obtained for the value $\theta_0$. By convention, we index by zero all quantities evaluated at $\theta_0$. It is clear that the parametrisation is one-to-one in the sense that

$$\theta \neq \theta_0 \quad \Longrightarrow \quad (\delta, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_q) \neq (\delta_0, \boldsymbol{\alpha}_{01}, \dots, \boldsymbol{\alpha}_{0p}, \boldsymbol{\beta}_{01}, \dots, \boldsymbol{\beta}_{0q}),$$

for $i = 1, \dots q$, and $j = 1, \dots p$. To avoid confusion with the *parameter* $\theta$ we refer to $\delta, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_q, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p$ as the *functional* parameters of the model. We assume that $\theta_0$ belongs to a compact subset $\Theta$ of $\mathbb{R}^{M+(p+q)M^2}$ which will be further specified later.

**Remark 2.** The implication of Eq. (3.2) is that the volatility process $(\sigma_t^2)_{t\in\mathbb{Z}}$ belongs to the $M$-dimensional subspace of $H$ spanned by $\varphi_1, \dots, \varphi_M$. This is also assumed in Aue et al. (2017). An alternative nonparametric approach will be developed in Section 4.

Our estimator is defined as follows:

$$\widehat{\theta}_n := \operatorname*{argmin}_{\theta\in\Theta} \widetilde{Q}_n(\theta), \tag{3.4}$$

where

$$\widetilde{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_t(\theta), \quad \tilde{\ell}_t(\theta) = \sum_{m=1}^{M} \left\{ \frac{\langle y_t^2, \varphi_m\rangle}{\langle \tilde{\sigma}_t^2, \varphi_m\rangle} + \log\langle \tilde{\sigma}_t^2, \varphi_m\rangle \right\}, \tag{3.5}$$

and where the empirical volatility $\tilde{\sigma}_t^2$ is computed recursively as

$$\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\theta) = \delta + \sum_{i=1}^{q} \boldsymbol{\alpha}_i(y_{t-i}^2) + \sum_{j=1}^{p} \boldsymbol{\beta}_j(\tilde{\sigma}_{t-j}^2), \quad \text{for } t = 1, \dots, n, \tag{3.6}$$

with some fixed initial values $y_0^2, \dots, y_{-q+1}^2$ and $\tilde{\sigma}_0^2, \dots, \tilde{\sigma}_{-p+1}^2$ in $H^+$. Note that the positivity of the baseline functions $\varphi_k$ ensures that the scalar products in (3.5) are positive and thus $\widetilde{Q}_n(\theta)$ is well defined and reaches its minimum on the compact set $\Theta$. This estimator is clearly inspired by the QMLE for standard GARCH models, and thus, we will refer to $\widehat{\theta}_n$ as the QML estimator.

### 3.2. Asymptotic results

Under (3.2), Model (2.1)–(2.2) admits a multivariate representation. More precisely, under the invertibility Assumption **A5** below, we define the $\mathbb{R}^M$-valued process $(h_t(\theta))_{t\in\mathbb{Z}}$ as the stationary and ergodic solution of the following equation

$$h_t(\theta) = \mathfrak{d} + \sum_{i=1}^{q} \mathfrak{A}_i Y_{t-i}^{\langle 2\rangle} + \sum_{j=1}^{p} \mathfrak{B}_j h_{t-j}(\theta), \tag{3.7}$$

where $Y_t^{\langle 2\rangle} = (\langle y_t^2, \varphi_1\rangle, \dots \langle y_t^2, \varphi_M\rangle)'$, $\mathfrak{d} = \Phi d$ and for $i = 1, \dots, q$ and $j = 1, \dots, p$, $\mathfrak{A}_i = \Phi A_i$ and $\mathfrak{B}_j = \Phi B_j$ with $\Phi = (\langle \varphi_i, \varphi_j\rangle)$ being the Gram-matrix of the functions $\varphi_1, \dots, \varphi_M$. Note that $h_t(\theta_0) = (\langle \sigma_t^2, \varphi_1\rangle, \dots, \langle \sigma_t^2, \varphi_M\rangle)'$.

We are able to deduce our main asymptotic results under the following assumptions:

**A1** $\theta_0 \in \Theta$, $\Theta$ is a compact set.

**A2** $E\|\eta_0^2\|_\infty^\tau < \infty$ for some $\tau \in (0, 1)$, $(y_t)_{t\in\mathbb{Z}}$ is a strictly stationary and non-anticipative solution of Model (2.1)–(2.2).

**A3** For any function $\psi \in H$ and any non-random constant $\kappa$,

$$\langle \eta_t^2, \psi\rangle = \kappa \ \text{ a.s.} \quad \Rightarrow \quad \psi \equiv 0 \ \text{ and } \ \kappa = 0.$$

**A4** If $p > 0$, $\mathfrak{A}_0(z) = \sum_{i=1}^{q} \Phi A_{0i} z^i$ and $\mathfrak{B}_0(z) = I_M - \sum_{j=1}^{p} \Phi B_{0i} z^i$, are left co-primes and $[A_{0q}, B_{0p}]$ has full rank $M$.

**A5** $\delta(u) > 0$ for all $u$, and $\mathfrak{A}_i$ and $\mathfrak{B}_j$ have non-negative entries for all $\theta \in \Theta$. Furthermore the matrix $\mathfrak{B}(z) = I_M - \sum_{j=1}^{p} \Phi B_j z^i$ is invertible for $|z| \le 1$.

[1] Note that $K_{\alpha_i}(u, v) = \sum_{k,\ell=1}^{M} a_{k,\ell}^{(i)} \varphi_k(u)\varphi_\ell(v)$ and $K_{\beta_i}(u, v) = \sum_{k,\ell=1}^{M} b_{k,\ell}^{(i)} \varphi_k(u)\varphi_\ell(v)$.

**Theorem 2.** *Under (3.1)–(3.2) and Assumptions* **A1–A5**, *the QMLE of $\theta_0$ is strongly consistent, i.e. we have $\widehat{\theta}_n \to \theta_0$ almost surely.*

In order to derive the asymptotic law of our estimator we make the following assumptions:

**A6** $\theta_0 \in \mathrm{Int}(\Theta)$.

**A7** $E\|\eta_0\|_\infty^4 < \infty$.

**Remark 3.** Assumptions **A4**, **A5**, **A6** and **A7** are the functional analogues to the standard regularity assumptions for CCC-GARCH processes, see e.g. in Francq and Zakoïan (2012). Assumption **A2** is needed to apply Proposition 2. Unlike in the multivariate setting, the first part of **A2** is not implied by the moment condition in (3.1). Indeed, consider the following counterexample: $\eta_0(u) = M\mathbb{1}_{u=U} + \mathbb{1}_{u\neq U}$, where $M$ is a random variable with only logarithmic moments and $U$ is a uniform variable over [0, 1]. Assumption **A3** means that the distribution of $\eta_t^2$ as a random function is non-degenerate. For instance, suppose that $\eta_t(u) = \eta_t(u+c)$ for all $u \in [a, b]$ and some $c \in (0, 1-b)$, then we see that the condition fails by choosing e.g. $\psi = \mathbb{1}_{[a,b]} - \mathbb{1}_{[a+c,b+c]}$. The first part of Assumption **A5** is satisfied if $d$, $A_i$ and $B_i$ have non-negative entries. Finally, note that the last condition in **A5** ensures that the initial values in (3.6) do not matter asymptotically.[2]

Letting $\mathcal{N}_K(\mu, \Sigma)$ denote a $K$-variate normal random vector with mean $\mu$ and covariance matrix $\Sigma$, we get the following asymptotic normality result. Define $\ell_t(\theta)$ by replacing $\langle\tilde{\sigma}_t^2, \varphi_m\rangle$ by $h_{t,m}$ (the $m$-th component of $h_t(\theta)$) in $\tilde{\ell}_t(\theta)$.

**Theorem 3.** *Under (3.1)–(3.2) and Assumptions* **A1–A7** *we have that*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}_{M+(p+q)M^2}\left(0, J^{-1}IJ^{-1}\right), \tag{3.8}$$

*where* $I = Var\left(\frac{\partial \ell_t(\theta_0)}{\partial\theta}\right)$ *and*

$$J = E\left[\frac{\partial^2 \ell_t(\theta_0)}{\partial\theta\partial\theta'}\right] = \sum_{m=1}^{M} E\left[\frac{1}{h_{t,m}^2}\frac{\partial h_{t,m}}{\partial\theta}\frac{\partial h_{t,m}}{\partial\theta'}(\theta_0)\right]. \tag{3.9}$$

We remark that, unlike in the scalar case, it is not possible to factorise the matrix $I$ in the asymptotic variance of Theorem 3. However, we have that

$$I = \sum_{m,m'=1}^{M} E\left[\left(\iint E[\eta_t^2(u)\eta_t^2(v)]\frac{\sigma_t^2(u)\sigma_t^2(v)}{h_{t,m}(\theta_0)h_{t,m'}(\theta_0)}\varphi_m(u)\varphi_{m'}(v)dudv - 1\right)\right.$$
$$\left.\times \frac{1}{h_{t,m}h_{t,m'}}\frac{\partial h_{t,m}}{\partial\theta}\frac{\partial h_{t,m'}}{\partial\theta'}(\theta_0)\right].$$

One can thus estimate $J$ and $I$ by

$$\widehat{J} = \frac{1}{n}\sum_{m=1}^{M}\sum_{t=1}^{n}\left[\frac{1}{\tilde{h}_{t,m}(\widehat{\theta}_n)\tilde{h}_{t,m'}(\widehat{\theta}_n)}\frac{\partial\tilde{h}_{t,m}(\widehat{\theta}_n)}{\partial\theta}\frac{\partial\tilde{h}_{t,m'}(\widehat{\theta}_n)}{\partial\theta'}\right]$$

and

$$\widehat{I} = \frac{1}{n}\sum_{m,m'=1}^{M}\sum_{t=1}^{n}\left[\left(\iint \widehat{K}_{\eta_0^2}(u, v)\frac{\tilde{\sigma}_t^2(\widehat{\theta}_n)(u)\,\tilde{\sigma}_t^2(\widehat{\theta}_n)(v)}{\tilde{h}_{t,m}(\widehat{\theta}_n)\tilde{h}_{t,m'}(\widehat{\theta}_n)}\varphi_m(u)\varphi_{m'}(v)dudv - 1\right)\right.$$
$$\left.\times \frac{1}{\tilde{h}_{t,m}(\widehat{\theta}_n)\tilde{h}_{t,m'}(\widehat{\theta}_n)}\frac{\partial\tilde{h}_{t,m}(\widehat{\theta}_n)}{\partial\theta}\frac{\partial\tilde{h}_{t,m'}(\widehat{\theta}_n)}{\partial\theta'}\right],$$

where the vector $\tilde{h}_t(\widehat{\theta}_n)$ and its derivatives are computed recursively by using Eq. (3.7),

$$\tilde{\sigma}_t^2(\widehat{\theta}_n)(u) = \sum_{m=1}^{M}(\Phi^{-1}\tilde{h}_t(\widehat{\theta}_n))_m\varphi_m(u),$$

and

$$\widehat{K}_{\eta_0^2}(u, v) = \frac{1}{n}\sum_{t=1}^{n} y_t^2(u)/\tilde{\sigma}_t^2(\widehat{\theta}_n)(u)\cdot y_t^2(v)/\tilde{\sigma}_t^2(\widehat{\theta}_n)(v).$$

2 In our numerical illustrations, we took the mean of the squares on the first week $\sum_{t=1}^{5} y_t^2$ for the initial values $y_{-i}^2$ and $\tilde{\sigma}_{-i}^2$.

### 3.3. Specification of the baseline functions

The functions $\varphi_1, \ldots, \varphi_M$ must satisfy positivity constraints. We can consider any family of non-negative and linearly independent functions in $H$, such as the power basis $1, u, u^2, \ldots$ the exponential basis $e^u, e^{2u}, \ldots$ or some polynomial basis that is non-negative on $[0, 1]$, for example the popular B-splines bases. Since the latter are thought to perform well with functional data, we will consider them in our empirical study. More precisely, we will use the Bernstein polynomials. For more details on the use of B-splines and smoothing methods for functional data see Ramsay and Silverman (2005).

Alternatively, we propose a data-driven method to get the baseline functions. Direct use of functional PCA (which is by far the most common practice in many applications) is not possible in this framework, due to the positivity constraints. Functional principal components define orthonormal functions, and as a matter of fact, they cannot be jointly positive. Under the further assumption that $E\|y_t\|^4 < \infty$, we represent the squared process through its functional principal components, i.e.

$$y_t^2(u) = \mu(u) + \sum_{j=1}^{\infty} \langle y_t^2 - \mu, \psi_j \rangle \psi_j(u),$$

where $\mu(u) = E[y_t^2(u)]$ and $(\psi_j)_{j\geq 1}$ are the eigenfunctions of the covariance operator of $y_t^2$ (see, e.g. Horváth and Kokoszka, 2014 for more details on functional principal components). Then, since $\sigma_t^2 = E[y_t^2|\mathcal{F}_{t-1}]$ and in view of (2.2), it seems natural to assume that $\delta, \boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ are spanned by the finite set of functions $\mu, \psi_1, \ldots, \psi_{M-2}$. Since these functions are not positive we propose to modify them according to the following routine. Take $\varphi_1(u) = 1$, $\varphi_2(u) = \mu(u)$, which is necessarily non-negative, and shift the other principal components, if necessary:

$$\varphi_m(u) := \psi_{m-2}(u) - \inf_{u\in[0,1]} \psi_{m-2}(u) \wedge 0, \quad \text{for all } m = 3, \ldots, M. \tag{3.10}$$

In practice the eigenfunctions and the mean function are replaced by their respective empirical versions. We have observed in our simulations (see Section 5.1.2) that this empirical choice performs relatively well, even when we compare it to the settings where the true (but unknown) basis functions in the data-generating process were used.

To validate the choice of the number $M$ and more generally the correct specification of the model, we suggest three different strategies:

- **Static approach:** Since $E[y_t^2] = E[\sigma_t^2] \in \mathrm{sp}\{\varphi_1, \ldots, \varphi_M\}$. We can check on a plot if $\frac{1}{n}\sum_{t=1}^n y_t^2$ is well approximated by $M$ basis functions.
- **ARMA representation:** From (3.7) at $\theta_0$ we can deduce the ARMA-type equation

$$Y_t^{\langle 2\rangle} = \mathfrak{d}_0 + \sum_{i=1}^{p\vee q} (\mathfrak{A}_{0i} + \mathfrak{B}_{0i}) Y_{t-i}^{\langle 2\rangle} - \sum_{j=1}^{p} \mathfrak{B}_{0j} u_{t-j},$$

  where $Y_t^{\langle 2\rangle} = (\langle y_t^2, \varphi_1\rangle, \ldots \langle y_t^2, \varphi_M\rangle)'$ and $u_t = Y_t^{\langle 2\rangle} - h_t(\theta_0)$. Although $(u_t)_{t\in\mathbb{Z}}$ is not an i.i.d. sequence, it can be shown to be weak white noise under some moment assumptions. There exist validation tests for *weak* vector ARMA models that could be used to assess the choice of the basis functions, see Boubacar Maïnassara and Saussereau (2018).
- **Out of sample comparison:** We can measure the quality of prediction of the squared returns by computing the *empirical integrated loss* as

$$\widehat{\mathrm{EIL}}_M = \sum_{t=n_0+1}^{n-1} \int L\big(\widetilde{\sigma}_{t+1}^2(\widehat{\theta}_{1:t}^{(M)})(u),\, y_{t+1}^2(u)\big)\, du,$$

  where $L$ is some loss function and $\widehat{\theta}_{1:t}^{(M)}$ denotes the QML estimator based on the sub-sample $y_1, \ldots, y_t$ and the basis functions $\varphi_1, \ldots, \varphi_M$. The choice of $M$ and the basis functions is based on the minimisation of $\widehat{\mathrm{EIL}}_M$. For the loss function $L$, we can take the mean square error $L_{\mathrm{LS}}(x, y) = |x - y|^2$ or the `Qlike` $L_{\mathtt{Qlike}}(x, y) = x/y + \log(y)$, see Patton (2011). This backtesting method can be seen as a particular case of cross-validation adapted to time series. We have displayed in Figs. 2 and 3 the values of $\widehat{\mathrm{EIL}}_M$ for $M = 2, \ldots 6$ with $n_0 = n - 10$, computed from the simulated process defined below in Section 5.1, Example 2. We observe that $L_{\mathtt{Qlike}}$ performs significantly better than $L_{\mathrm{LS}}$ since, in this setting, the true value is $M = 3$.

## 4. Extension to infinite-dimensional parameter space

Assuming a finite-dimensional parametrisation (3.2) may appear unsatisfactory from the theoretical standpoint. In this section, we show that the QML estimator remains strongly consistent in a more general setting, permitting an infinite-dimensional specification. For simplicity, we only consider the case when $p = q = 1$. We assume that $\delta$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be parametrised by some infinite-dimensional parameter $\theta \in \Theta$, i.e. we no longer assume (3.2). This parameter space is assumed to be a compact subset of $l^2$ (the set of square summable sequences).

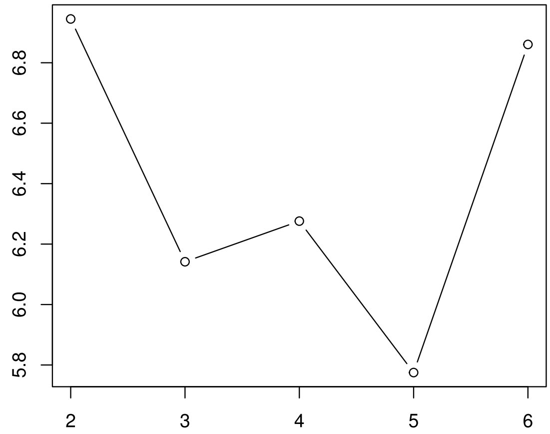


**Fig. 2.** $\widehat{\mathrm{EIL}}_M$ for $L = L_{\mathrm{LS}}$.

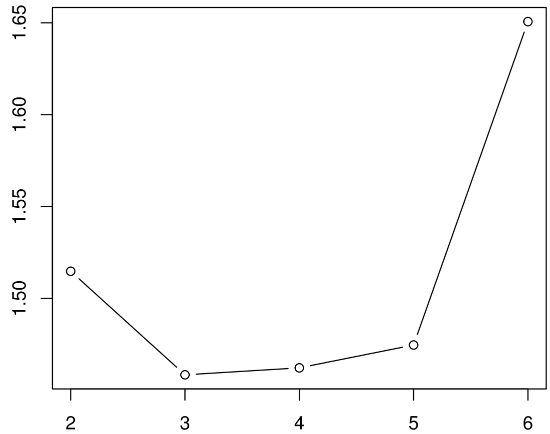


**Fig. 3.** $\widehat{\mathrm{EIL}}_M$ for $L = L_{\mathtt{Qlike}}$.

Let $\varphi_1, \varphi_2, \ldots$ be non-negative and linearly independent functions in $H$. Our new estimator is defined as

$$\widehat{\theta}_n^N := \underset{\theta \in \Theta_N}{\operatorname{argmin}}\, \widetilde{Q}_n(\theta),$$

where $\Theta_N \subset \Theta$ is the subspace of all sequences with zero entries in components $k > N$. Furthermore, $\widetilde{Q}_n(\theta)$ is defined as in (3.5) with

$$\widetilde{\ell}_t(\theta) = \sum_{m=1}^{\infty} w_m \left\{ \frac{\langle y_t^2, \varphi_m \rangle}{\langle \tilde{\sigma}_t^2(\theta), \varphi_m \rangle} + \log \langle \tilde{\sigma}_t^2(\theta), \varphi_m \rangle \right\}, \tag{4.1}$$

where $(w_m)_{m \geq 1}$ is a positive and summable sequence of numerical weights.

Let $\boldsymbol{\alpha}^*$ denote the *adjoint* operator of $\boldsymbol{\alpha}$, i.e. the unique operator such that $\langle \boldsymbol{\alpha}^*(x), y \rangle = \langle x, \boldsymbol{\alpha}(y) \rangle$ for all $x, y \in H$. The following technical assumptions will be used.

**A8** For all $\theta \in \Theta$, we have

(a) $\langle \delta_0, \varphi_m \rangle = \langle \delta, \varphi_m \rangle$ for all $m \geq 1$ implies $\delta = \delta_0$.
(b) $\boldsymbol{\alpha}_0^*(\varphi_m) = \boldsymbol{\alpha}^*(\varphi_m)$ for all $m \geq 1$ implies $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$.
(c) $(\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}_0^*)(\varphi_m) = (\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}^*)(\varphi_m)$ for all $m \geq 1$ implies $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

**A9** There exists a positive sequence $(a_m)_{m \geq 1}$ such that $w_m / a_m^2$ is summable and $a_m \leq \int \varphi_m(u) du \leq \|\varphi_m\| \leq 1$ for all $m \geq 1$. Furthermore, for all $\theta \in \Theta$, the function $\delta$ is uniformly bounded from below by some constant $c > 0$.

**A10** $\delta(u) > 0$ for all $u$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are non-negative operators and $\|\boldsymbol{\beta}\| < 1$ for all $\theta \in \Theta$.

**Remark 4.** Note that under (3.2) (i.e. when the parameters have finite rank) with $p = q = 1$, it can be shown that if $\mathfrak{A}_0$ is invertible, then **A8** is satisfied. Assumption **A9** is a technical assumption that is needed for (i) and (iii) in the proof of Proposition 3. As for the weights $(w_m)$, any positive sequence satisfying $\sum_{m\geq 1} w_m/a_m^2 < \infty$ can be used, where $0 < a_m < \int_0^1 \varphi_m(u)du$. For example, if $\varphi_m(u) = u^{m-1}$, then $\int_0^1 \varphi_m(u)du = 1/m$, which means that $w_m$ has to decay at a faster rate than $m^{-3}$. The sequence thus depends on the choice of the baseline functions. Finally, **A10** is the analogue of assumption **A5**.

**Proposition 3.** *Under (3.1) and assumptions* **A1**–**A3**, *and* **A8**–**A10**, *the QMLE of $\theta_0$ is strongly consistent: $\widehat{\theta}_n^{N_n} \to \theta_0$ almost surely in $l^2$, for any sequence $N_n \nearrow \infty$.*

To illustrate this result, we present an example of a functional GARCH process which is not included in the multivariate and linearly parametrised setting of Section 3.

Let $(\psi_k)_{k\geq 1}$ be an orthonormal basis of $H$. We assume that the volatility dynamics is of the form

$$\sigma_t^2(u) = \exp\left(\sum_{k=1}^{\infty} d_k \psi_k(u)\right) + a\int y_{t-1}^2(v)dv + b\int \sigma_{t-1}^2(v)dv. \tag{4.2}$$

Here, the parameter is $\theta = (a, b, d_1, d_2, \dots) \in \mathbb{R}_+^2 \times \mathbb{R}^\infty$. The exponential function is used to guarantee a positive intercept, but other positive valued functions could be used instead. This model provides a very simple interpretation of the function $\delta$. Indeed, we can show that the curve of $\delta$ parallels that of the expected intraday-volatility. More precisely, if $Ey_t^2(u) < \infty$, then we have that

$$E\sigma_t^2(u) = Ey_t^2(u) = \delta(u) + \frac{a+b}{1-a-b}\int \delta(v)dv.$$

We can also compute explicitly the top Lyapunov exponent which only depends on $a$, $b$ and the law of $\eta_0$:

$$\begin{aligned}\gamma &= \lim_{t\to\infty}\frac{1}{t}E\log\left\|(\boldsymbol{\alpha}\Upsilon_{t-1}+\boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\Upsilon_1+\boldsymbol{\beta})\right\|\\ &= \lim_{t\to\infty}\frac{1}{t}E\log\sup_{\|x\|\leq 1}\prod_{s=2}^{t-1}(a\int\eta_s^2(v)dv+b)\int(a\eta_1^2(v)+b)x(v)dv\\ &= E\log\left(a\int\eta_0^2(v)dv+b\right).\end{aligned}$$

Finally, Proposition 3 can be applied to model (4.2). Indeed, we can easily choose a compact subset $\Theta$ of $l^2$ and an innovation process $(\eta_t)_{t\in\mathbb{Z}}$ such that Assumptions **A1**−**A3** are satisfied. If we assume that $(d_k)_{k\geq 1}$ is absolutely summable and that $\kappa = \sup_{k\geq 1}\|\psi_k\|_\infty < \infty$, then we get that $\delta(u) = \exp\left(\sum_{k=1}^\infty d_k\psi_k(u)\right) \geq \exp\left(-\kappa\sum_{k=1}^\infty |d_k|\right) > 0$ and assumption **A10** is satisfied provided that $b < 1$. Now, since $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, it is easy to see that **A8** (b) is satisfied as well as **A8** (c), provided that $a_0 \neq 0$. It is not very restrictive to further assume that $\delta$ belongs to the subspace of $H$ spanned by the functions $\varphi_1, \varphi_2, \dots$ and, thus, that **A8** (a) holds for the chosen sequence $(\varphi_m)_{m\geq 1}$. Then, for any weight sequence $(w_m)_{m\geq 1}$ which satisfies **A9** we get by Proposition 3 that the QMLE of model (4.2) is strongly consistent.

## 5. Numerical results

In this section we first compare our estimation approach with other existing methods on different setups. Then we illustrate the potential usefulness of the functional GARCH on real data examples.

### 5.1. Simulations

The estimator of Aue et al. (2017) is defined as follows

$$\widehat{\theta}_n^{\mathrm{LS}} = \operatorname*{argmin}_{\theta\in\Theta}\widetilde{Q}_n(\theta), \quad \text{where} \quad \widetilde{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^n\sum_{m=1}^M(\langle y_t^2, \phi_m\rangle - \langle\tilde{\sigma}_t^2(\theta), \phi_m\rangle)^2$$

where $\phi_1, \dots, \phi_M$ are orthonormal functions in $H$ (without the positivity constraint). We will compare this Least Squares Estimator (`LSE`) and our `QMLE` that is defined in (3.4). We will next compare the QML with given baseline functions $\varphi_m$ to the data-driven procedure described in Section 3.3 (we then refer to `QMLE*`). In order that the parameter space $\Theta$ satisfies Assumptions **A1** and **A5** we will impose that $0 \leq b_{k\ell} \leq .99 \cdot (M^2 \cdot \max_{1\leq m\leq M}\|\varphi_m\|)^{-1}$, for all $k, \ell = 1, \dots, M$.

**Table 1**
Performance of LSE.

| | $d$ | $a$ | $b$ |
|---|---|---|---|
| sd | 0.013 | 0.096 | 0.3 |
| Bias | 0.087 | 0.066 | 0.08 |

**Table 2**
Performance of QMLE.

| | $d$ | $a$ | $b$ |
|---|---|---|---|
| sd | 0.0026 | 0.069 | 0.099 |
| Bias | 0.082 | 0.016 | 0.012 |

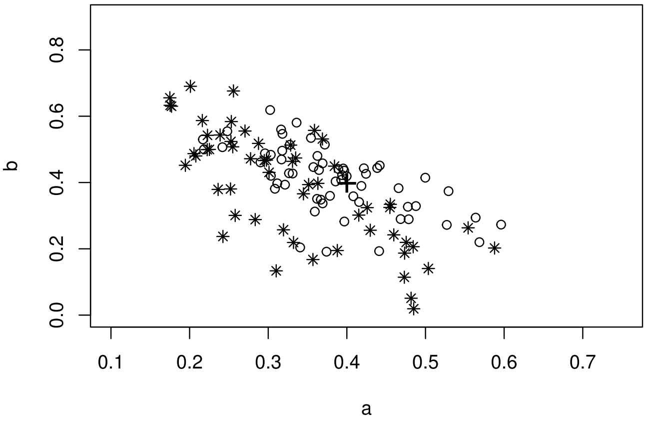


**Fig. 4.** 60 estimates of $a$ and $b$, with $*$ LSE and $\circ$ QMLE, and $+$ indicates the true values.

#### 5.1.1. Example 1

The first setup is taken from Aue et al. (2017). They consider a GARCH(1,1) model with

$$\delta(u) = .01, \; K_\alpha(u, v) = K_\beta(u, v) = 12u(1-u)v(1-v), \tag{5.1}$$

for $u, v \in [0, 1]$. For the innovations, Ornstein–Uhlenbeck processes are chosen. They are defined as $\eta_0(u) = e^{-u/2}W_0(e^u)$, where $(W_0(u))_{u\in[0,1]}$ is a Brownian motion.

The recursion starts at initial value $\sigma_0^2 := \delta$, and the first 1000 curves are discarded. Aue et al. (2017) project on one basis function $\varphi_1(u) = \sqrt{30}u(1-u)$, $u \in [0, 1]$. It follows that $K_\alpha(u, v) = a\,\varphi_1(u)\varphi_1(v)$, with $a = 0.4$ and $K_\beta(u, v) = b\,\varphi_1(u)\varphi_1(v)$, with $b = 0.4$. Note that $\delta$ is not spanned by $\varphi_1(u)$ and that $d = \langle \delta, \varphi_1 \rangle \approx .009$. It is assumed that $\varphi_1$ is known and we estimate $d$, $a$ and $b$. For the LSE we impose $|b| \leq .99$, whereas, for the QMLE we impose that $a \geq 0$ and $0 \leq b \leq .99$. In order to compare the performance of the two procedures, we consider 1000 Monte-Carlo replications of our estimation experiment for a sample size $n = 600$. The results of our simulations are displayed in Tables 1–2 and in Fig. 4. We see that standard deviation and bias differ by a factor of 2 to 3 in favour of the QMLE method, which is confirmed by Fig. 4.

#### 5.1.2. Example 2

We now illustrate our estimator in a slightly more complex example. We consider a functional GARCH(1,1) model with $\delta(u) = (u - .5)^2 + .1$,

$$K_\alpha(u, v) = (u - .5)^2 + (v - .5)^2 + .2, \quad \text{and} \quad K_\beta(u, v) = (u - .5)^2 + (v - .5)^2 + .4. \tag{5.2}$$

As in the previous example we take for the innovations an i.i.d. sequence of Ornstein–Uhlenbeck processes. The recursion starts at initial value $\sigma_0^2 := \delta$, and the first 1000 curves were discarded. Simulated trajectories of this process are displayed in Fig. 1.

For the baseline functions $\varphi_1, \ldots, \varphi_M$ we consider the following families:

1. QMLE: Bernstein polynomials, which are special cases of B-spline functions defined by $\varphi_k(u) = \binom{M-1}{k-1} u^{k-1}(1-u)^{M-k}$, for $k = 1, \ldots, M$ and $u \in [0, 1]$.
2. QMLE*: The functions defined in Section 3.3.
3. LSE: The orthogonal polynomials $\varphi_1(u) = 1$, $\varphi_2(u) = u^2 - 1/2$ and $\varphi_3(u) = u^2 - u + 5/6$.
4. LSE*: Bernstein polynomials, i.e. the same as for the QMLE.

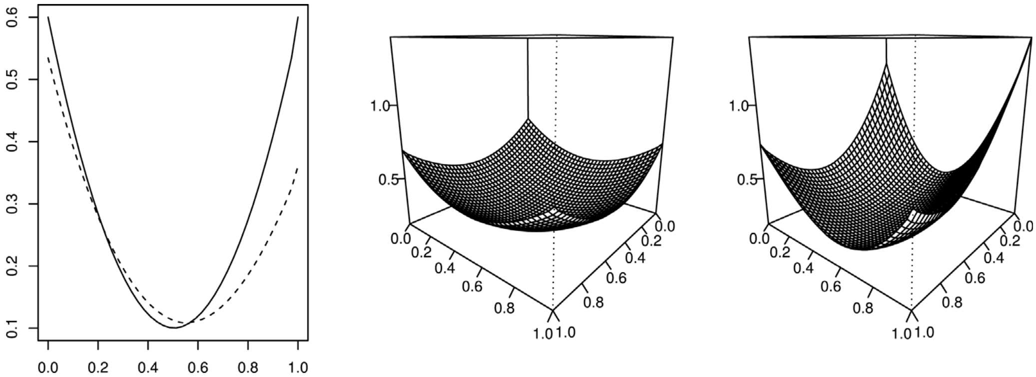


**Fig. 5.** From left to right: the intercept function $\delta$ (solid line) compared to its estimation $\hat{\delta}$ (dashed line), the theoretical kernel $K_{\boldsymbol{\alpha}}(u, v)$ and the estimated kernel $K_{\widehat{\boldsymbol{\alpha}}}(u, v)$.

**Table 3**
Relative mean squared deviations for the functional parameters.

| | $\delta$ | $\boldsymbol{\alpha}$ | $\boldsymbol{\beta}$ |
|---|---|---|---|
| QMLE | 0.43 | 0.4 | 0.49 |
| QMLE* | 0.58 | 0.66 | 0.78 |
| LSE | 1.2 | 0.46 | 0.6 |
| LSE* | 1.6 | 0.94 | 1.3 |

**Table 4**
Relative mean squared deviations for the functional parameters.

| | $\delta$ | $\boldsymbol{\alpha}$ |
|---|---|---|
| QMLE | 0.31 | 0.33 |
| ME | 0.46 | 0.99 |

By fixing $M = 3$, the subspace spanned by the Bernstein polynomials (of order 2) contains the true parameters defined in (5.2). We have constrained the parameters as follows: $d_k \geq 10^{-5}$, $a_{k\ell} \geq 0$ and $0 \leq b_{k\ell} \leq .99 \cdot (M^2 \cdot \max_{1 \leq m \leq M} \|\varphi_m\|)^{-1}$, for all $k, \ell = 1, \ldots, M$.

The functional parameters $\delta$ and $\boldsymbol{\alpha}$ (its kernel $K_{\boldsymbol{\alpha}}$) are represented in Fig. 5, together with one particular realisation of the QMLE. In order to compare the performance of the three different procedures, we ran $N = 100$ Monte-Carlo replications of our estimation experiment with sample size $n = 1000$. The results are displayed in Table 3. We show the relative mean squared deviations, defined for $\delta$ and $\boldsymbol{\alpha}$ by

$$\frac{1}{N^{1/2}\|\delta\|}\left(\sum_{v=1}^{N}\|\widehat{\delta}^{(v)} - \delta\|^2\right)^{1/2}, \quad \frac{1}{N^{1/2}\|\boldsymbol{\alpha}\|}\left(\sum_{v=1}^{N}\|\widehat{\boldsymbol{\alpha}}^{(v)} - \boldsymbol{\alpha}\|^2\right)^{1/2},$$

and analogously for $\boldsymbol{\beta}$. As already mentioned in Section 3.3, it is interesting to note that both procedures perform similarly, despite the fact that QMLE* does not require prior knowledge of baseline functions.

### 5.1.3. Example 3

Finally, we performed a simulation with the same setting as in Example 2 except that $K_{\boldsymbol{\beta}}(u, v) = 0$ for all $u$ and $v$, i.e. we consider a functional ARCH(1) model. We can thus compare our QMLE to the moment estimator (ME) proposed by Hörmann et al. (2013). Their estimator is defined as

$$K_{\widehat{\boldsymbol{\alpha}}}(u, v) = \frac{1}{n-1}\sum_{t=1}^{n}\sum_{j=1}^{K}\sum_{i=1}^{K}\hat{\lambda}_j^{-1}\langle y_t, \hat{e}_j\rangle\langle y_{t+1}, \hat{e}_i\rangle\hat{e}_j(v)\hat{e}_i(u) \quad \text{and} \quad \widehat{\delta} = (\mathrm{I}_H - \widehat{\boldsymbol{\alpha}})\frac{1}{n}\sum_{t=1}^{n} y_t^2,$$

where $(\hat{\lambda}_j)_{j\geq 1}$ and $(\hat{e}_j)_{j\geq 1}$ are respectively, the empirical eigenvalues and eigenfunctions of the covariance operator of $y_t^2$ and $K$ is a tuning parameter. As suggested by the authors, we take the largest integer such that $\hat{\lambda}_K/\hat{\lambda}_1 \geq .01$, which results here in $K = 6$. The results of our simulations displayed in Table 4 show that, once again, the QMLE outperforms its competitor.

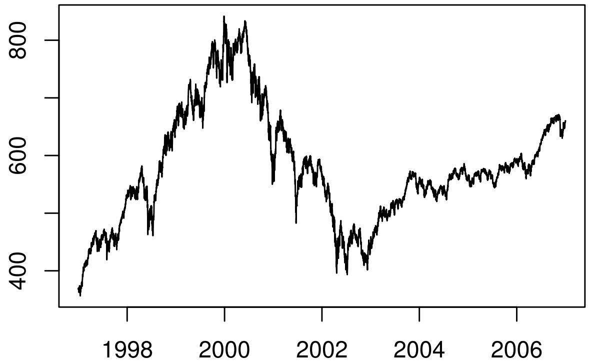


**Fig. 6.** Raw data for S&P 100 index between 1997 and 2007.

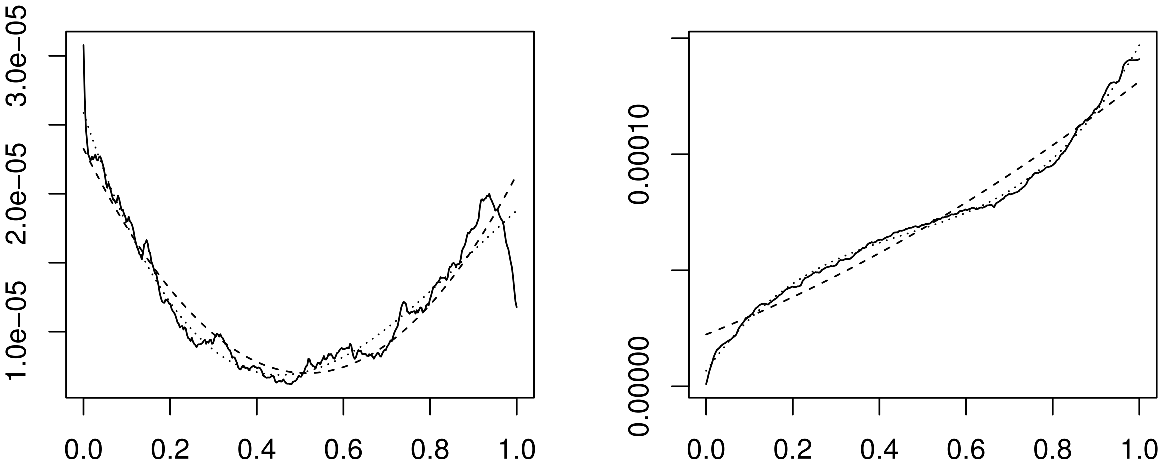


**Fig. 7.** The solid lines represent the empirical means of $y_t^2$ (left panel) and $\tilde{y}_t{}^2$ (right panel), the dashed and dotted lines represent their projections on $\varphi_1, \ldots, \varphi_3$ and $\varphi_1, \ldots, \varphi_4$, respectively.

### 5.2. Real data illustration

We applied our estimators to the minutely recorded S&P100 Index for a ten-year period between 1997 and 2007. The return series is displayed in Fig. 6.

The functional GARCH model has been implemented on two different types of return data. Denoting by $X_t(u)$ the price at time $u$ of the day $t$, we considered the $\tau$-minute returns $y_t(u)$ and the intraday returns $\tilde{y}_t(u)$ defined by

$$y_t(u) = \log X_t(u) - \log X_t(u - \tau), \quad \text{and} \quad \tilde{y}_t(u) = \log X_t(u) - \log X_t(0). \tag{5.3}$$

For $y_t(u)$ we used $\tau = 20$ min.[3] For the baseline functions we used, as in Section 5.1.2 the Bernstein polynomials and choose $M = 4$. This choice is motivated by the *static approach* proposed in Section 3.3. Indeed, Fig. 7 shows that the empirical means of $y_t^2$ and $\tilde{y}_t{}^2$ are much better approximated (except on the boundaries) by a linear combination of 4 rather than 3 Bernstein polynomials. We computed the LS estimator to get an initial value of the parameter in the optimisation routine. The estimated kernels for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are displayed in Fig. 8 for $y_t$ and $\tilde{y}_t$. The top panels reveal that the volatility of $y_t(u)$ decreases from the beginning ($u$ close to 0) to the end of the day ($u$ close to 1), that the effect of $y_{t-1}^2(v)$ is more important for $v$ close to 0 (beginning of the day) or 1 (end of the day), and that the effect of $\sigma_{t-1}^2(v)$ is roughly uniform when $v$ varies. Similar comments hold for $\tilde{y}_t$ (bottom panels) except that the volatility tends to increase within the day. This is not surprising since, by construction, $\tilde{y}_t(u)$ has similarities with a random walk when $u$ increases from 0 to 1 (more precisely, $\tilde{y}_t(K\tau) = \sum_{s=1}^{K} y_t(s\tau)$, where $K$ increases from 1 to $\lfloor 1/\tau \rfloor$). The empirical volatility curves are displayed in Figs. 9 and 10 for $y_t$ and in Figs. 11 and 12 for $\tilde{y}_t$. In light of (2.3) and the related discussion, we plotted the curves $\tilde{\sigma}_t(\widehat{\theta}_n)(u) \cdot \widehat{Q}_{1-\alpha/2}^{\hat{\eta}(u)}$. The required quantiles were estimated from the residuals $\hat{\eta}_t(u) := y_t(u)/\tilde{\sigma}_t(\widehat{\theta}_n)(u)$, for $t = 1, \ldots, n$. On both processes, we

[3] Note that the functional approach is not very appropriate when $\tau$ is less than, say 10 min, since then, the curves become very irregular and, thus, there is no chance to capture the intraday dynamics by smoothing the signal. As noted by the Associate Editor, another reason for not considering $\tau$ too small is the presence of the microstructure noise.

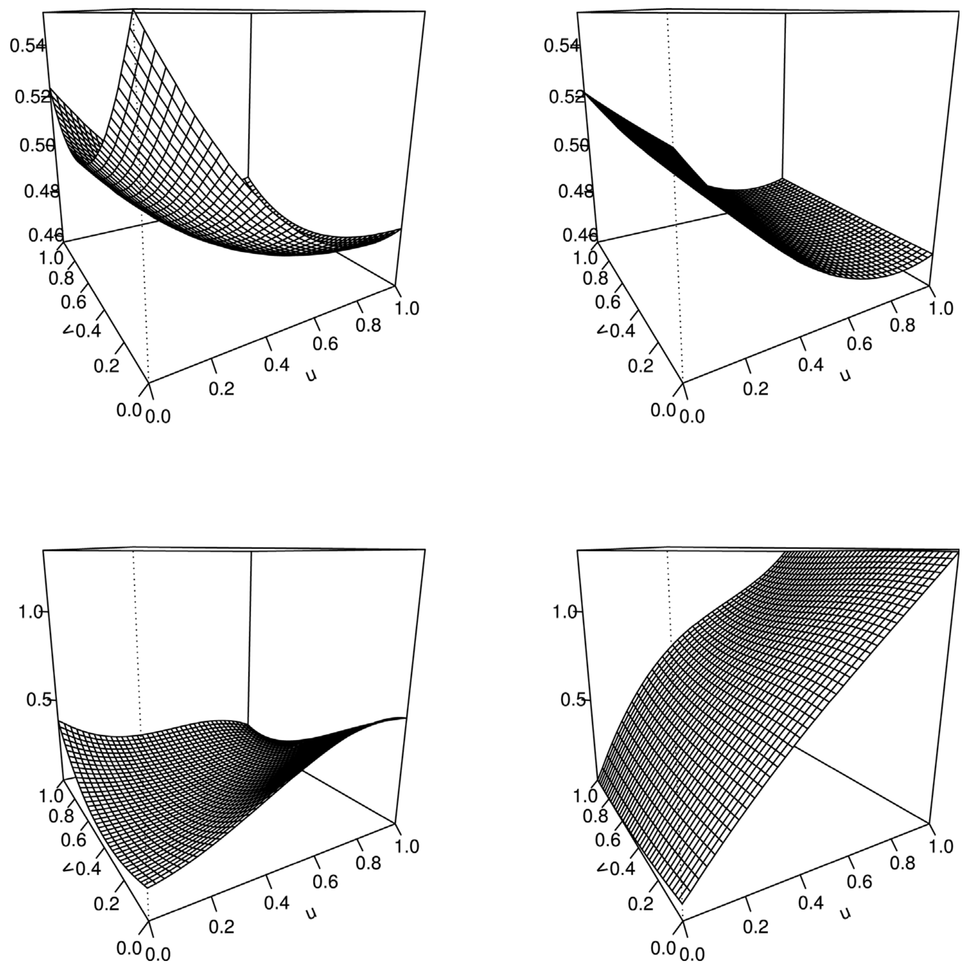


**Fig. 8.** The estimated kernels $K_{\widehat{\alpha}}(u, v)$ (left panels) and $K_{\widehat{\beta}}(u, v)$ (right panels), for $y_t$ (top panels) and $\tilde{y}_t$ (bottom panels).

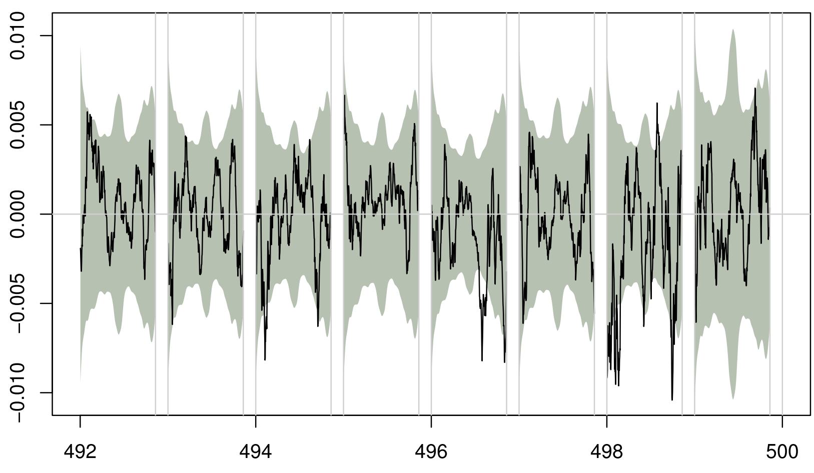


**Fig. 9.** Predicted volatility (shaded area) for $y_t$ (8 days).

can observe the sensitivity to shocks of the volatility process and its persistence. The persistence seems stronger in Fig. 12 than in Fig. 10, whereas the rise of the volatility after a shock is more evident in Fig. 9 than in Fig. 11. This is in line with a large value of $\|\widehat{\boldsymbol{\beta}}\|$ and a small value of $\|\widehat{\boldsymbol{\alpha}}\|$ for $\tilde{y}_t$ (see Table 5).

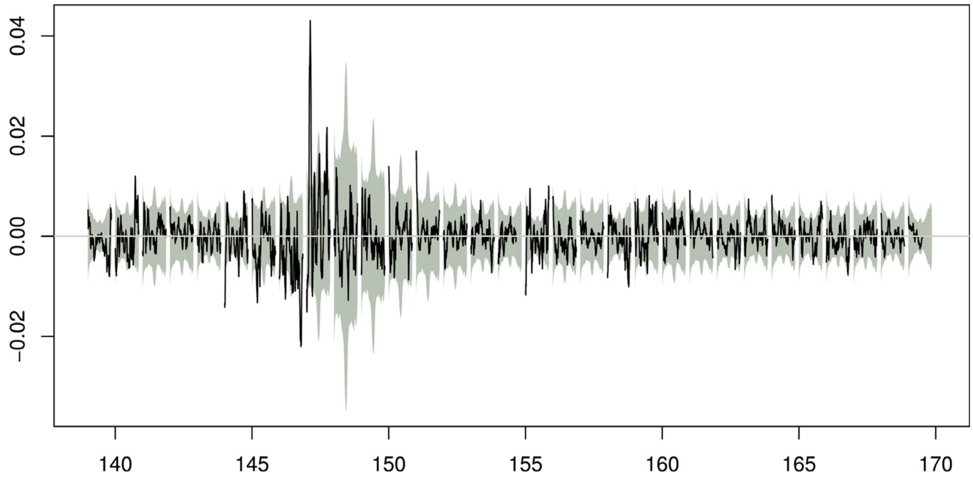


**Fig. 10.** Predicted volatility (shaded area) for $y_t$ (31 days).

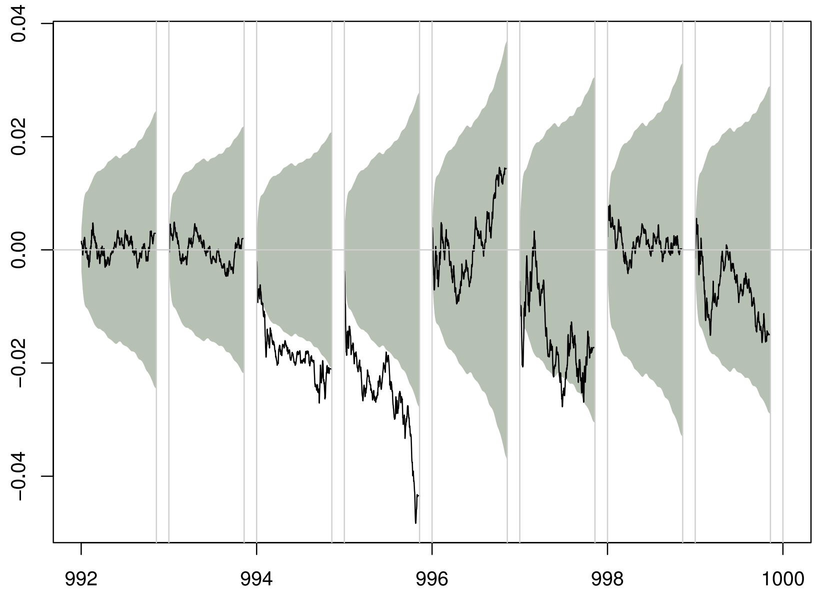


**Fig. 11.** Predicted volatility (shaded area) for $\tilde{y}_t$ (8 days).

**Table 5**
Norms of the estimated parameters.

| | $\Vert\hat{\delta}\Vert$ | $\Vert\hat{\boldsymbol{\alpha}}\Vert$ | $\Vert\hat{\boldsymbol{\beta}}\Vert$ |
|---|---|---|---|
| $y$ | 1e−06 | 0.46 | 0.46 |
| $\tilde{y}$ | 5e−06 | 0.15 | 0.89 |

*Realised volatility*

Practitioners often use the so-called *realised volatility* as a measure of the daily risk. Typically, it is defined as follows:

$$\mathrm{RV}_t = \sum_{j=1}^{\lfloor 1/\tau \rfloor} |\log X_t(j\tau) - \log X_t(j\tau - \tau)|^2. \tag{5.4}$$

If we choose the same $\tau$ as in the definition of $y_t$ in (5.3) we remark that

$$\mathrm{RV}_t = \sum_{j=1}^{\lfloor 1/\tau \rfloor} |y_t^2(j\tau)| = \sum_{j=1}^{\lfloor 1/\tau \rfloor} \sigma_t^2(j\tau)\eta_t^2(j\tau).$$

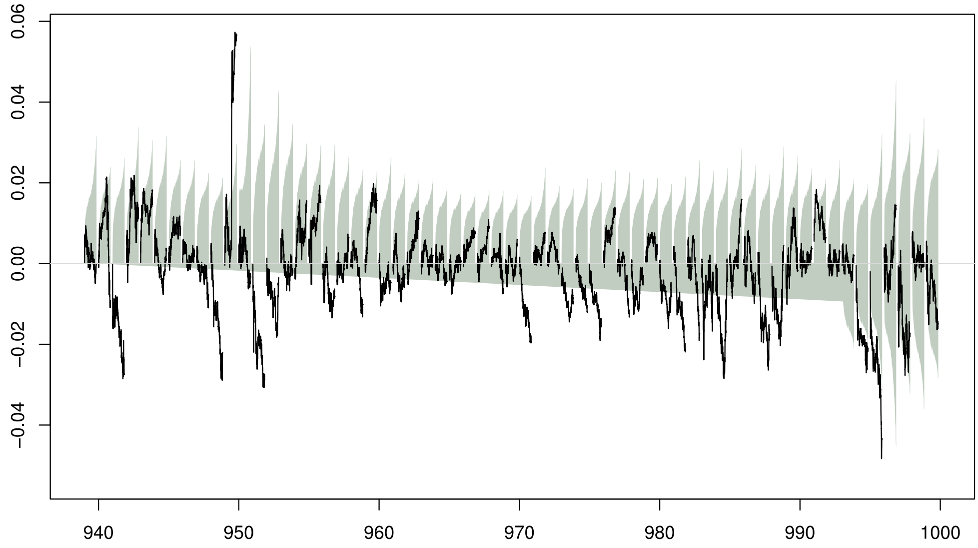

**Fig. 12.** Predicted volatility (shaded area) for $\tilde{y}_t$ (61 days).

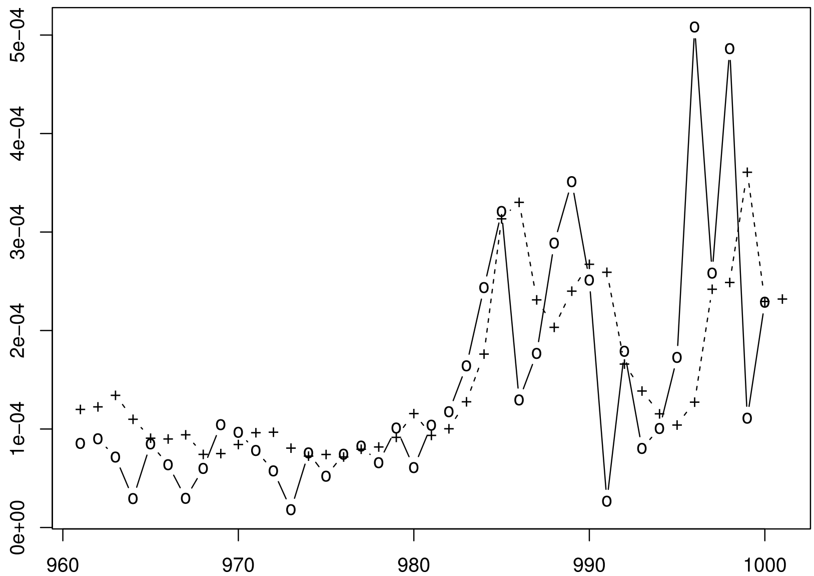

**Fig. 13.** Predicted (+) and true realised volatility (∘).

At time $t-1$, the optimal predictor of $\mathrm{RV}_t$ is

$$E[\mathrm{RV}_t|\mathcal{F}_{t-1}] = \sum_{j=1}^{\lfloor 1/\tau \rfloor} \sigma_t^2(j\tau),$$

which can be estimated by

$$\widetilde{\mathrm{RV}}_t = \sum_{j=1}^{\lfloor 1/\tau \rfloor} \widetilde{\sigma}_t^2(\widehat{\theta})(j\tau), \tag{5.5}$$

where $\widehat{\theta}$ is the QMLE computed with the sub-sample $y_1, \dots, y_{t-1}$. In Fig. 13, we have plotted 41 one-day ahead predictions of $\widetilde{\mathrm{RV}}_t$ against $\mathrm{RV}_t$. Similar predictions could be obtained using more standard models, for instance using the HAR model of Corsi (2009). The interest of our approach is that other realised volatility measures could be handled using the same estimated model.

## 6. Conclusion

This paper provides stationarity conditions and develops a QML-type estimator for a general class of functional GARCH($p, q$) models. Conditions for the consistency and asymptotic normality were established in cases where the estimation reduces to a finite dimensional problem. Our numerical illustrations showed that these results/models can be used to analyse intra-day financial data.

We also proved consistency of the QML estimator in the infinite-dimensional case, but the asymptotic normality remains an open issue. Indeed, it is generally difficult to prove weak convergence in infinite dimensional spaces. To get some idea of the difficulties, consider the so-called function-on-scalar regression

$$Y = \langle \rho, X \rangle + \epsilon, \tag{6.1}$$

where $\epsilon$ and $X$ are independent and $\rho$ is the slope—which here is a function in $L^2$. Estimation of $\rho$ from an i.i.d. sample $(Y_i, X_i)_{1\le i\le n}$ has been broadly investigated. But even in the relatively simple setting of (6.1), Cardot et al. (2007) proved that it is impossible to obtain a CLT (in the topology of the considered functional space) for a large class of estimators $\hat{\rho}$. Of course, such a negative result cannot be directly translated in our context. We leave the investigations concerning the asymptotic distribution in the infinite-dimensional for future research.

## Acknowledgements

We are most thankful to the Co-Editor, Jianqing Fan, and two anonymous referees for insightful comments and suggestions. This research has been partly supported by Communauté francaise de Belgique, Actions de Recherche Concertées, Projects Consolidation 2016–2021. The first author acknowledges support from the F.R.S.-FNRS Fonds de la Recherche Scientifique, Q9 15 Rue d'Egmont 5, B-1000 Bruxelles. The second and fourth authors are grateful to the Agence Nationale de la Recherche (ANR), which supported this work via the Project MultiRisk (ANR CE26 2016 - CR), and the Labex ECODEC. The fourth author acknowledges financial support from the Research Initiative "Quantitative Management Research Initiative (QMI)" under the aegis of the Fondation du Risque a joint initiative by University Paris-Dauphine, ENSAE ParisTech, Addstones-GFI and La Française Investment Solutions.

## Appendix

We start by showing asymptotic results for the semi-strong CCC-GARCH models, which will be used to prove Theorems 2 and 3. The model was introduced by Bollerslev (1990) and further studied under the assumption of i.i.d. innovations, among others, by Aue et al. (2009) and Francq and Zakoïan (2012).

### *A.1. Asymptotics of semi-strong CCC-GARCH models*

We recall the definition of a CCC-GARCH($p, q$) process. It is an $\mathbb{R}^M$-valued process $(\epsilon_t)_{t\in\mathbb{Z}}$ with $\epsilon_t = (\epsilon_{t,1}, \ldots, \epsilon_{t,M})$ which satisfies the following equations:

$$\epsilon_t = H_t^{1/2}\nu_t, \tag{A.1}$$

$$H_t = D_t R D_t \quad \text{with} \quad D_t = \big(\mathrm{diag}(h_t)\big)^{1/2}, \tag{A.2}$$

$$h_t = \mathfrak{d} + \sum_{i=1}^{q} \mathfrak{A}_i \epsilon_{t-i}^{[2]} + \sum_{j=1}^{p} \mathfrak{B}_j h_{t-j}, \tag{A.3}$$

where $\epsilon_t^{[2]} = (\epsilon_{t,1}^2, \ldots, \epsilon_{t,M}^2)$, $(\eta_t)_{t\in\mathbb{Z}}$ is a sequence of innovations and $R$ is an $M \times M$ correlation matrix, $\mathfrak{A}_i$ and $\mathfrak{B}_j$ are $M \times M$ matrices with non-negative elements, the components of the $M$-vector $\mathfrak{d}$ are strictly positive. Note that the correlation matrix $R$ has to be estimated as well.

In general this model is studied under the assumption of i.i.d. innovations, but this hypothesis is too strong in our framework. We therefore consider a *semi-strong* version of the CCC-GARCH, under Assumption **A*3** below.[4]

We set

$$\xi = \big(\mathrm{vec}'\big(\mathfrak{d}, \mathfrak{A}_1, \ldots, \mathfrak{A}_q, \mathfrak{B}_1, \ldots, \mathfrak{B}_p\big), r'\big)',$$

where $r$ is the vector of the subdiagonal elements of $R$. The QMLE $\hat{\xi}_n$ of $\xi_0$ is defined by

$$\hat{\xi}_n = \arg\min_{\xi\in\Xi} \frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_t(\xi), \qquad \tilde{\ell}_t(\xi) = \epsilon_t'\tilde{H}_t^{-1}\epsilon_t + \log|\det(\tilde{H}_t)|, \tag{A.4}$$

where $\tilde{H}_t$ is defined recursively using (A.3) and some initial values.

[4] The convergence and asymptotic properties of semi-strong GARCH models has been considered by Escanciano (2009) in the scalar case. We provide a multivariate generalisation of this result to semi-strong CCC-GARCH models.

We define the matrix-valued polynomials $\mathfrak{A}(z) = \sum_{i=1}^{q} \mathfrak{A}_i z^i$ and $\mathfrak{B}(z) = I_M - \sum_{j=1}^{p} \mathfrak{B}_j z^j$, for $z \in \mathbb{C}$ and any $\xi$ belonging to a compact parameter set $\Xi$. Let $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ be some filtration. The following technical assumptions are needed:

**A**$^*$**0** $E\|\epsilon_t^{[2]}\|^s < \infty$, for some $s > 0$.

**A**$^*$**1** $\xi_0 \in \Xi$, where $\Xi$ is a compact set.

**A**$^*$**2** $(\epsilon_t)$ is a strictly stationary and ergodic solution of Model (A.1)–(A.3), with $\epsilon_t \in \mathcal{F}_t$.

**A**$^*$**3** $(\nu_t)_{t\in\mathbb{Z}}$ is an ergodic, stationary martingale difference sequence with respect to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ such that $E[\nu_t\nu_t'|\mathcal{F}_{t-1}] = I_M$. There exists no vector $x \neq 0 \in \mathbb{R}^M$ such that $x'\epsilon_t^{[2]}$ is $\mathcal{F}_{t-1}$-measurable.

**A**$^*$**4** If $q > 0$, $\mathfrak{A}_0(z)$ and $\mathfrak{B}_0(z)$, are left co-primes and $[\mathfrak{A}_{0q}, \mathfrak{B}_{0p}]$ has full rank $M$.

**A**$^*$**5** $\inf_{\xi\in\Xi} \mathfrak{d} > 0$ componentwise; $\mathfrak{B}(z)$ is invertible for $|z| \leq 1$, for all $\xi \in \Xi$; $R$ is a positive definite correlation matrix for all $\xi \in \Xi$.

**A**$^*$**6** $\xi_0 \in \mathrm{Int}(\Xi)$.

**A**$^*$**7** $E\|\nu_t\nu_t'\|^{2(1+w)} < \infty$, for some $w > 0$.

Let $\ell_t(\xi) = \epsilon_t' H_t^{-1}(\xi)\epsilon_t + \log|\det(H_t(\xi))|$.

**Theorem 4.** *Under Assumptions* **A**$^*$**0**–**A**$^*$**5** *the QMLE of $\xi_0$ as defined in (A.4) is strongly consistent, i.e.* $\widehat{\xi}_n \to \xi_0$ *a.s.*

**Theorem 5.** *Under Assumptions* **A**$^*$**0**–**A**$^*$**7**, *we have that*

$$\sqrt{n}(\hat{\xi}_n - \xi_0) \xrightarrow{d} \mathcal{N}_{M+(p+q)M^2+M(M-1)/2}\left(0, J^{-1}IJ^{-1}\right) \tag{A.5}$$

*where* $I = Var\left(\frac{\partial \ell_t(\xi_0)}{\partial \xi}\right)$ *and* $J = E\left[\frac{\partial^2 \ell_t(\xi_0)}{\partial\xi\partial\xi'}\right]$.

**Proof of Theorem 4.** A close look into the proof of Theorem 11.7 in Francq and Zakoïan (2011) shows that independence of the innovations is only needed to show the existence of some small order moments and for the identifiability. The existence of moments is now imposed in **A**$^*$**0**. To show the identifiability, we have to prove that if there exists a matrix $P_1 \in \mathbb{R}^{M\times M}$ such that $P_1\epsilon_t^{[2]} = Z_{t-1}$, *a.s.* where the vector $Z_{t-1}$ is $\mathcal{F}_{t-1}$ measurable, then $P_1 = 0$. Using the second part of our assumption **A**$^*$**3** we may conclude. □

**Proof of Theorem 5.** As one can see in the proof of Theorem 11.8 in Francq and Zakoïan (2011), the independence of the innovations is only used to show the existence of moments of the theoretical criterion $\ell_t$, or of its derivative at $\xi_0$ or at a neighbourhood of it. For example, in order to prove the existence of $I$, we first compute for $i \leq s := M + (p+q)M^2$,

$$\begin{aligned}\frac{\partial \ell_t(\xi_0)}{\partial \xi_i} &= -\mathrm{Tr}\left\{(\epsilon_t\epsilon_t' D_t^{-1}R^{-1} + R^{-1}D_t^{-1}\epsilon_t\epsilon_t')D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_i}D_t^{-1}\right\} + 2\mathrm{Tr}\left\{D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_i}\right\}\\ &= \mathrm{Tr}\left\{(I_M - R^{-1/2}\nu_t\nu_t'R^{1/2})\frac{\partial D_t(\xi_0)}{\partial \xi_i}D_t^{-1} + (I_M - R^{1/2}\nu_t\nu_t'R^{-1/2})D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_i}\right\}.\end{aligned}$$

When independence between $\nu_t$ and the past holds, the existence of the second-order moments of these derivatives only requires $E\|\nu_t\|^4 < \infty$. Under our martingale difference assumption **A**$^*$**3**, the moment condition on $\nu_t$ has to be strengthened as in **A**$^*$**7**. More precisely, for $i, j \leq s$, we use Hölder's inequality to get that

$$\begin{aligned}E\left|\frac{\partial \ell_t(\xi_0)}{\partial \xi_i}\frac{\partial \ell_t(\xi_0)}{\partial \xi_j}\right| &\leq \mathrm{const}\cdot\left(1 + E\|\nu_t\nu_t'\|^{2(1+w)}\right)^{\frac{1}{1+w}}\left(E\left\|D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_i}\frac{\partial D_t(\xi_0)}{\partial \xi_j}D_t^{-1}\right\|^{\frac{1+w}{w}}\right)^{\frac{w}{1+w}}\\ &\leq \mathrm{const}\cdot\left(E\left\|D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_i}\right\|^{\frac{2(1+w)}{w}} E\left\|D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_j}\right\|^{\frac{2(1+w)}{w}}\right)^{\frac{w}{2(1+w)}} < \infty.\end{aligned}$$

The case $i > s$, i.e. when deriving with respect to the coefficients of the matrix $R$, is actually much simpler. The remainder of the proof works as in the classical CCC-GARCH by similar arguments as in Francq and Zakoïan (2011). □

*A.2. Proofs of the results of Sections 2 and 4*

**Proof of Theorem 1.** By iterating (2.4), we formally get that

$$\underline{z}_t = \underline{b}_t + \sum_{k=1}^{\infty} \boldsymbol{\Psi}_t \boldsymbol{\Psi}_{t-1} \cdots \boldsymbol{\Psi}_{t-k+1}(\underline{b}_{t-k}). \tag{A.6}$$

The series converges almost surely, since using (2.8), we deduce that

$$\limsup_{t\to\infty} \frac{1}{t} \log \|\boldsymbol{\Psi}_t \boldsymbol{\Psi}_{t-1} \cdots \boldsymbol{\Psi}_{t-k+1}(\underline{b}_{t-k})\| \leq \gamma + \limsup_{t\to\infty} \frac{1}{t} \log \|\underline{b}_{t-k}\|, \quad \text{a.s.} \tag{A.7}$$

Since $E \log^+ \|\underline{b}_{t-k}\| < \infty$ by (2.5), and $\|\underline{b}_{t-k}\| \geq \|\delta\| > 0$ the second summand in the right hand side of (A.7) is zero and, thus, we can apply the Cauchy rule to show convergence. In addition, it is easy to see that the $q+1$-th component of $\underline{z}_t$ defines a non-anticipative and stationary solution of (2.2). The proof of the existence is complete.

It remains to prove that the solution is almost surely unique. To this end, let us assume that $\tilde{\underline{z}}_t^*$ is another solution. By iterating (2.4), we get that

$$\underline{z}_t^* = \underline{z}_t^N + \boldsymbol{\Psi}_t \cdots \boldsymbol{\Psi}_{t-N}(\underline{z}_{t-N-1}^*), \quad \text{where } \underline{z}_t^N = \underline{b}_t + \sum_{k=1}^{N} \boldsymbol{\Psi}_t \cdots \boldsymbol{\Psi}_{t-k+1}(\underline{b}_{t-k}).$$

We then deduce that

$$\|\underline{z}_t^* - \underline{z}_t\| \leq \|\underline{z}_t^N - \underline{z}_t\| + \|\boldsymbol{\Psi}_t \boldsymbol{\Psi}_{t-1} \cdots \boldsymbol{\Psi}_{t-N}\| \cdot \|\underline{z}_{t-N-1}^*\|. \tag{A.8}$$

We already know that since $\gamma < 0$, we have $\|\underline{z}_t^N - \underline{z}_t\| \to 0$ and $\|\boldsymbol{\Psi}_t \cdots \boldsymbol{\Psi}_{t-N}\| \to 0$, almost surely when $N \to \infty$. Furthermore, the law of $\|\underline{z}_{t-N-1}^*\|$ is independent of $N$. Hence, the right-hand side of (A.8) tends to zero in probability. Therefore, $P(\underline{z}_t^* = \underline{z}_t) = 1$. □

**Proof of Corollary 1.** First, note that

$$\begin{aligned}
\boldsymbol{\Psi}_t \boldsymbol{\Psi}_{t-1} \cdots \boldsymbol{\Psi}_1 &= \begin{bmatrix} \boldsymbol{\Upsilon}_t \\ \boldsymbol{I}_H \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Upsilon}_{t-1} \\ \boldsymbol{I}_H \end{bmatrix} \cdots \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Upsilon}_1 \\ \boldsymbol{I}_H \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\Upsilon}_t \\ \boldsymbol{I}_H \end{bmatrix} (\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{t-1} + \boldsymbol{\beta}) \cdots (\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1 + \boldsymbol{\beta}) \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \end{bmatrix},
\end{aligned}$$

from which we can deduce a bound for the top Lyapunov exponent:

$$\begin{aligned}
\gamma &\leq \lim_{t\to\infty} \frac{1}{t} \left\{ E \log(\left\|\eta_t^2\right\|_\infty + 1)^{1/2} + E \log \left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{t-1} + \boldsymbol{\beta}) \cdots (\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1 + \boldsymbol{\beta})\right\| + E \log \left\| \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \end{bmatrix} \right\| \right\} \\
&= \lim_{t\to\infty} \frac{1}{t} E \log \left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{t-1} + \boldsymbol{\beta}) \cdots (\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1 + \boldsymbol{\beta})\right\| \\
&= \inf_{t\geq 1} \frac{1}{t} E \log \left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_t + \boldsymbol{\beta}) \cdots (\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1 + \boldsymbol{\beta})\right\|.
\end{aligned}$$

The first inequality is in fact an equality since the two side terms are vanishing in the limit and the last equality follows from the Fekete's lemma. □

**Proof of Proposition 1.** Since $\gamma < 0$, we deduce from Theorem 1 that there exists a strictly stationary and non-anticipative solution $(y_t)_{t\in\mathbb{Z}}$ to (2.1)–(2.2). Using stationarity and independence we deduce

$$E\underline{z}_t = E\underline{b}_t + \sum_{k=1}^{\infty} (E\boldsymbol{\Psi}_0)^k E\underline{b}_0, \tag{A.9}$$

which converges, since by assumption $\rho(E\boldsymbol{\Psi}_0) < 1$. This implies that $E[\sigma_t^2(u)] < \infty$ and $E[y_t^2(u)] < \infty$ for all $u \in [0, 1]$. □

**Proof of Proposition 2.** Using (2.7), there exists an integer $t_0$ such that $E \log \|\boldsymbol{\Psi}_{t_0} \boldsymbol{\Psi}_{t_0-1} \cdots \boldsymbol{\Psi}_1\|^\tau < 0$. Furthermore, we have that

$$E\|\boldsymbol{\Psi}_{t_0} \boldsymbol{\Psi}_{t_0-1} \cdots \boldsymbol{\Psi}_1\|^\tau \leq E\|\boldsymbol{\Psi}_{t_0}\|^\tau \|\boldsymbol{\Psi}_{t_0-1}\|^\tau \cdots \|\boldsymbol{\Psi}_1\|^\tau = (E\|\boldsymbol{\Psi}_1\|^\tau)^{t_0},$$

where we used the fact that $(\boldsymbol{\Psi}_t)_{t\in\mathbb{Z}}$ are i.i.d. in the last equality. Note that $E\|\boldsymbol{\Psi}_1\|^\tau < \infty$ by (2.6). From Lemma 2.2 in Francq and Zakoïan (2011), we then deduce that there exists an $0 < s < \tau$ such that $\varsigma := E\|\boldsymbol{\Psi}_{t_0}\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_1\|^s < 1$. From (A.6) we get that

$$\begin{aligned}
E\|\underline{z}_t\|^s &\le E\|\underline{b}_0\|^s\left\{1+\sum_{k=1}^{\infty}E\|\boldsymbol{\Psi}_k\boldsymbol{\Psi}_{k-1}\cdots\boldsymbol{\Psi}_1\|^s\right\}\\
&\le E\|\underline{b}_0\|^s\left\{1+\sum_{\ell=0}^{\infty}\sum_{r=0}^{t_0-1}E\|\boldsymbol{\Psi}_k\cdots\boldsymbol{\Psi}_{k-r+1}\|^s E\|\boldsymbol{\Psi}_{\ell t_0}\cdots\boldsymbol{\Psi}_{(\ell-1)t_0+1}\|^s\ldots E\|\boldsymbol{\Psi}_{t_0}\cdots\boldsymbol{\Psi}_1\|^s\right\}\\
&\le E\|\underline{b}_0\|^s\left\{1+\sum_{\ell=0}^{\infty}\varsigma^\ell\sum_{r=0}^{t_0-1}\left(E\|\boldsymbol{\Psi}_1\|^s\right)^r\right\},
\end{aligned}$$

where by convention $E\|\boldsymbol{\Psi}_{\ell t_0}\cdots\boldsymbol{\Psi}_{(\ell-1)t_0+1}\| = 1$ for $\ell = 0$ and $E\|\boldsymbol{\Psi}_k\cdots\boldsymbol{\Psi}_{k-r+1}\| = 1$ for $r = 0$. Furthermore, we have that

$$E\|\underline{b}_0\|^s \le E\|\eta_0^2\delta\|^s + \|\delta\|^s \le \left(E\|\eta_0^2\|_\infty^s + 1\right)\|\delta\|^s < \infty.$$

Thus $E\|\underline{z}_t\|^s < \infty$ and the conclusion follows. □

### A.3. Proofs of the results of Section 3

In view of representation (3.7) at $\theta_0$, we build a sequence $(\epsilon_t)$ satisfying the CCC-GARCH model of Section 3. Let $(r_t)_{t\in\mathbb{Z}}$ be an i.i.d. sequence of $M$-dimensional vectors, whose components are independent Rademacher variables. Let $\epsilon_t = \{\mathrm{diag}(Y_t^{\langle 2\rangle})\}^{1/2}r_t$ and let $\mathcal{F}_t$ the $\sigma$-field generated by $(r_t, \{\eta_{t-u}, u \ge 0\})$. Let $D_t = \{\mathrm{diag}(\langle\sigma_t^2, \varphi_m\rangle)\}^{1/2}$ and let $\nu_t = D_t^{-1}\epsilon_t$. Note that $\epsilon_t^{[2]} = (\epsilon_{t,1}^2, \ldots, \epsilon_{t,M}^2) = Y_t^{\langle 2\rangle}$. It follows that Eqs. (A.1)–(A.3) hold with $R = I_M$. Let

$$\xi = \mathrm{vec}\big(\mathfrak{d}, \mathfrak{A}_1, \ldots, \mathfrak{A}_q, \mathfrak{B}_1, \ldots, \mathfrak{B}_p\big) = (I_{1+M(p+q)} \otimes \Phi)\theta,$$

where $\otimes$ denotes the usual Kronecker product of matrices. Since $\Phi$ is non-singular (this follows from the linear independence of the functions $\varphi_1, \ldots, \varphi_M$), the transformation $T$ which maps $\theta$ to $\xi$ is bijective. By choosing $\Xi = T(\Theta)$, the QML estimator $\hat\xi_n$ defined in (A.4) satisfies $\hat\xi_n = T(\hat\theta_n)$. Clearly, Theorems 4–5 can be straightforwardly adapted when $R = I_M$ is not estimated. It therefore suffices to verify that the assumptions of these theorems are satisfied.

**Proof of Theorem 2.** We start by verifying that the multivariate process $(\epsilon_t)_{t\in\mathbb{Z}}$ defined by $\epsilon_t = \{\mathrm{diag}(Y_t^{\langle 2\rangle})\}^{1/2}r_t$ satisfies assumptions **A**$^*$**0**–**A**$^*$**5**.

Assumption **A**$^*$**0** follows from Proposition 2, noting that

$$\|\epsilon_t^{[2]}\| = \left(\sum_{m=1}^M |\langle y_t^2, \varphi_m\rangle|^2\right)^{1/2} \le \|y_t^2\|\left(\sum_{m=1}^M \|\varphi_m\|^2\right)^{1/2},$$

where $\|.\|$ denotes the euclidean norm. Assumption **A**$^*$**1** is obviously satisfied. By construction $\epsilon_t \in \mathcal{F}_t$ and satisfies (A.1)–(A.3). The stationarity and ergodicity of $\epsilon_t$ readily follows from **A2**. Thus **A**$^*$**2** holds. The first part of **A**$^*$**3** is obtained by noting that $E(\langle y_t^2, \varphi_m\rangle^{1/2}r_{t,m}|\mathcal{F}_{t-1}) = 0$ and $E(\langle y_t^2, \varphi_m\rangle|\mathcal{F}_{t-1}) = \langle\sigma_t^2, \varphi_m\rangle$. For the second part of **A**$^*$**3** we suppose that there exists an $x \in \mathbb{R}^M$, such that $x'\epsilon_t^{[2]}$ is $\mathcal{F}_{t-1}$-measurable. Then, conditionally on $\mathcal{F}_{t-1}$ we have that

$$x'\epsilon_t^{[2]} = \sum_{m=1}^M x_m\langle y_t^2, \varphi_m\rangle = \langle\eta_t^2, \sigma_t^2\sum_{m=1}^M x_m\varphi_m\rangle = \mathrm{const}, \quad \text{a.s.}$$

Assumption **A3** implies that the constant must be zero and that $\sigma_t^2(u)\sum_{m=1}^M x_m\varphi_m(u) = 0$, a.s. Furthermore, since $\sigma_t^2(u) \ge \delta_0(u) > 0$, for all $u \in [0, 1]$ and the function $\varphi_1, \ldots, \varphi_M$ are linearly independent, we can conclude that $x = 0$. The remaining assumptions **A**$^*$**4** and **A**$^*$**5** are obviously satisfied.

Hence, we can apply Theorem 4 to the process $(\epsilon_t)_{t\in\mathbb{Z}}$ and, thus, we get that $\widehat{\xi}_n \to \xi_0$, a.s. To conclude, the continuity of $T^{-1}$ implies that $\widehat{\theta}_n \to \theta_0$, a.s. □

**Proof of Theorem 3.** Since $T$ is a bijection, it is obvious that assumption **A6** implies that **A**$^*$**6**. Next, we have

$$\begin{aligned}
E\|\nu_t\nu_t'\|^2 &\le \sum_{k,m=1}^M E\nu_{t,k}^2\nu_{t,m}^2\\
&\le \sum_{k,m=1}^M \left(E\nu_{t,k}^4 E\nu_{t,m}^4\right)^{1/2}
\end{aligned}$$

$$= \sum_{k,m=1}^{M} \left(E\left[E[\langle \sigma_t^2\eta_t^2, \varphi_k\rangle^2 | \mathcal{F}_{t-1}]/\langle \sigma_t^2, \varphi_k\rangle^2\right] E\left[E[\langle \sigma_t^2\eta_t^2, \varphi_m\rangle^2 | \mathcal{F}_{t-1}]/\langle \sigma_t^2, \varphi_m\rangle^2\right]\right)^{1/2}$$
$$\leq M^2 E\|\eta_t\|_\infty^4,$$

where we used Hölder's inequality in the last step. By assumption **A7** we obtain that $E\|\nu_t\nu_t'\|^2 < \infty$. This is slightly weaker than assumption **A**$^*$**7**, which would require more than fourth order moments for the innovations process. However, in our situation we can avoid this further assumption and the use of Hölder's inequality as in Theorem 5. For example, to prove that

$$E\left\|\frac{\partial \ell_t(\theta_0)}{\partial\theta}\frac{\partial \ell_t(\theta_0)}{\partial\theta'}\right\| < \infty \quad \text{and} \quad E\left\|\frac{\partial^2 \ell_t(\theta_0)}{\partial\theta\partial\theta'}\right\| < \infty, \tag{A.10}$$

we compute

$$\frac{\partial \ell_t(\theta)}{\partial\theta} = \sum_{m=1}^{M}\left(1 - \frac{Y_{t,m}^{\langle 2\rangle}}{h_{t,m}}\right)\frac{1}{h_{t,m}}\frac{\partial h_{t,m}}{\partial\theta}, \tag{A.11}$$

$$\begin{aligned}\frac{\partial^2 \ell_t(\theta)}{\partial\theta\partial\theta'} &= \sum_{m=1}^{M}\left(1 - \frac{Y_{t,m}^{\langle 2\rangle}}{h_{t,m}}\right)\frac{1}{h_{t,m}}\frac{\partial^2 h_{t,m}}{\partial\theta\partial\theta'} \\ &\quad + \sum_{m=1}^{M}\left(2\frac{Y_{t,m}^{\langle 2\rangle}}{h_{t,m}} - 1\right)\frac{1}{h_{t,m}^2}\frac{\partial h_{t,m}}{\partial\theta}\frac{\partial h_{t,m}}{\partial\theta'}.\end{aligned} \tag{A.12}$$

Since, at the true value of the parameter $\theta = \theta_0$, we have that

$$\frac{Y_{t,m}^{\langle 2\rangle}}{h_{t,m_0}} = \frac{\langle y_t^2, \varphi_m\rangle}{\langle \sigma_t^2, \varphi_m\rangle} = \frac{\int \sigma_t^2(u)\eta_t^2(u)\varphi_m(u)du}{\int \sigma_t^2(u)\varphi_m(u)du} \leq \sup_{u\in[0,1]} \eta_t^2(u) = \|\eta_t^2\|_\infty. \tag{A.13}$$

This last quantity is independent of $\mathcal{F}_{t-1}$, which readily implies that (A.10) reduces to prove that

$$E\left|\frac{1}{h_{t,m}}\frac{\partial h_{t,m}}{\partial\theta_i}\right|^2 < \infty,$$

for all $m = 1, \ldots, M$ and $i = 1, \ldots, M + (p+q)M^2$. This can be established with Proposition 2. □

**Proof of Proposition 3.** We first prove that $\widehat{\theta}_n \to \theta_0$ almost surely, where $\widehat{\theta}_n$ denotes the minimiser of $\widetilde{Q}_n$ over the whole space $\Theta$. Let $\ell_t(\theta)$ denote the theoretical criterion (involving the infinite past of $y_t$), defined as $\tilde{\ell}_t(\theta)$, but with $\tilde{\sigma}_t^2(\theta)$ replaced by $\sigma_t^2(\theta)$ given by the model recursion. Note that under (3.1) we have that

$$E\left[\langle y_t^2, \varphi_m\rangle \mid \mathcal{F}_{t-1}\right] = \langle \sigma_t^2, \varphi_m\rangle. \tag{A.14}$$

To show $\widehat{\theta}_n \to \theta_0$ by standard arguments it suffices to verify the following:

(i) $\sup_{\theta\in\Theta} |Q_n(\theta) - \widetilde{Q}_n(\theta)| \underset{n\to\infty}{\longrightarrow} 0$ a.s., where $Q_n(\theta) = \frac{1}{n}\sum_{t=1}^n \ell_t(\theta)$;
(ii) $\langle \sigma_t^2(\theta), \varphi_m\rangle = \langle \sigma_t^2, \varphi_m\rangle$ a.s. for all $m \geq 1$ implies $\theta = \theta_0$;
(iii) $E\ell_t(\theta)$ exists for all $\theta \in \Theta$, and is finite for $\theta = \theta_0$, and $E\ell_t(\theta) > E\ell_t(\theta_0)$, for $\theta \neq \theta_0$;
(iv) $\forall\theta \neq \theta_0$, there exists a neighbourhood $\mathcal{V}_\theta$ such that $\liminf_n \inf_{\theta'\in\mathcal{V}_\theta} \widetilde{Q}_n(\theta') > E\ell_t(\theta_0)$, a.s.

Relation (iv) can be proven in the same way as in the univariate case. In order to prove (i), we recall that for all $\theta \in \Theta$, $\sigma_t^2(\theta) \geq \delta$. Hence, the non-negativity of the $\varphi_m$'s and **A9** implies that there is a strictly positive constant $c$, such that $\langle \sigma_t^2(\theta), \varphi_m\rangle \geq c a_m$ for all $m \geq 1$. Furthermore, we have that $\|\sigma_t^2(\theta) - \tilde{\sigma}_t^2(\theta)\| \leq K\rho^t$, where $\rho$ is the supremum of $\|\boldsymbol{\beta}_\theta\|$ over $\theta \in \Theta$ and $K$ is a random constant. The compactness of $\Theta$ and **A10** imply that $\rho < 1$. We then compute

$$\begin{aligned}&\sup_{\theta\in\Theta} |Q_n(\theta) - \widetilde{Q}_n(\theta)| \\ &\leq \frac{1}{n}\sum_{t=1}^{n}\sum_{m=1}^{\infty} w_m \sup_{\theta\in\Theta}\left\{\left|\frac{\langle \sigma_t^2(\theta), \varphi_m\rangle - \langle \tilde{\sigma}_t^2(\theta), \varphi_m\rangle}{\langle \tilde{\sigma}_t^2(\theta), \varphi_m\rangle\langle \sigma_t^2(\theta), \varphi_m\rangle}\right|\langle y_t^2, \varphi_m\rangle - \log\frac{\langle \tilde{\sigma}(\theta)_t^2, \varphi_m\rangle}{\langle \sigma_t^2(\theta), \varphi_m\rangle}\right\} \\ &\leq \sum_{m=1}^{\infty} w_m K\left(\sup_{\theta\in\Theta}\frac{1}{\langle \delta_\theta, \varphi_m\rangle^2}\right)\frac{1}{n}\sum_{t=1}^{n}\rho^t\langle y_t^2, \varphi_m\rangle + \sum_{m=1}^{\infty} w_m K\left(\sup_{\theta\in\Theta}\frac{1}{\langle \delta_\theta, \varphi_m\rangle}\right)\frac{1}{n}\sum_{i=1}^{n}\rho^t \\ &\leq K c^{-2}\sum_{m=1}^{\infty}\frac{w_m}{a_m^2}\frac{1}{n}\langle\sum_{t=1}^{n}\rho^t y_t^2, \varphi_m\rangle + K c^{-1}\sum_{m=1}^{\infty}\frac{w_m}{a_m}\frac{1}{n}\sum_{t=1}^{n}\rho^t.\end{aligned}$$

Consider the random function $Y = \lim_{n\to\infty} \sum_{t=1}^{n} \rho^t y_t^2$. The existence of moment of order $s$ for $y_t^2$ and its stationarity implies that $E\|Y\|^s \le \sum_{t=1}^{\infty} \rho^{ts} E\|y_t^2\|^s < \infty$, and, thus, that $Y$ is almost surely finite. We thus have that

$$\sup_{\theta\in\Theta} |Q_n(\theta) - \widetilde{Q}_n(\theta)| \le \frac{K\|Y\|}{nc^2} \sum_{m=1}^{\infty} \frac{w_m}{a_m^2} + \frac{K}{nc(1-\rho)} \sum_{m=1}^{\infty} \frac{w_m}{a_m} \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

We now turn to (iii). Although $\ell_t(\theta)$ is not necessarily integrable, it is well defined in $\mathbb{R} \cup \{\infty\}$, since by **A9**

$$E\ell_t(\theta) \ge E \sum_{m=1}^{\infty} w_m \log\langle \sigma_t^2(\theta), \varphi_m\rangle \ge \sum_{m=1}^{\infty} w_m(\log c + \log a_m) \\ \ge \text{const} - \sum_{m=1}^{\infty} w_m \log\Big(\frac{1}{a_m}\Big) > -\infty,$$

and at the true value of the parameter $\theta = \theta_0$, we have that

$$E\,\ell_t(\theta_0) = \sum_{m=1}^{\infty} w_m E\Big\{E\Big[\frac{\langle y_t^2, \varphi_m\rangle}{\langle \sigma_t^2, \varphi_m\rangle}\Big|\mathcal{F}_{t-1}\Big] + \log\langle \sigma_t^2, \varphi_m\rangle\Big\} \\ = \sum_{m=1}^{\infty} w_m\Big(1 + E\log\langle \sigma_t^2, \varphi_m\rangle\Big) \le \sum_{m=1}^{\infty} w_m\Big(1 + E\log\|\sigma_t^2\|\Big) < \infty.$$

We now have that

$$E[\ell_t(\theta)] - E[\ell_t(\theta_0)] = \sum_{m=1}^{\infty} w_m E\left\{\frac{\langle \sigma_t^2, \varphi_m\rangle}{\langle \sigma_t^2(\theta), \varphi_m\rangle} - 1 + \log\frac{\langle \sigma_t^2(\theta), \varphi_m\rangle}{\langle \sigma_t^2, \varphi_m\rangle}\right\} \\ \ge 0,$$

with equality if and only if for all $m \ge 1$

$$\langle \sigma_t^2(\theta), \varphi_m\rangle = \langle \sigma_t^2, \varphi_m\rangle, \quad \text{a.s.} \tag{A.15}$$

The proof of (iii) will be completed using (ii). To show (ii) we suppose that (A.15) holds true. We then have, for all $m \ge 1$,

$$\langle \delta_0, \varphi_m\rangle + \langle \boldsymbol{\alpha}_0(y_{t-1}^2), \varphi_m\rangle + \langle \boldsymbol{\beta}_0(\sigma_{t-1}^2), \varphi_m\rangle \\ = \langle \delta_\theta, \varphi_m\rangle + \langle \boldsymbol{\alpha}_\theta(y_{t-1}^2), \varphi_m\rangle + \langle \boldsymbol{\beta}_\theta(\sigma_{t-1}^2(\theta)), \varphi_m\rangle \quad \text{a.s.} \tag{A.16}$$

We have,

$$\langle \boldsymbol{\alpha}_\theta(y_{t-1}^2), \varphi_m\rangle = \langle \sigma_{t-1}^2\eta_{t-1}^2, \boldsymbol{\alpha}_\theta^*(\varphi_m)\rangle = \langle \eta_{t-1}^2, \sigma_{t-1}^2\boldsymbol{\alpha}_\theta^*(\varphi_m)\rangle.$$

In view of (A.16) we have

$$\langle \eta_{t-1}^2, \sigma_{t-1}^2[\boldsymbol{\alpha}_0^*(\varphi_m) - \boldsymbol{\alpha}_\theta^*(\varphi_m)]\rangle = K_{t-2} \quad \text{a.s.},$$

where $K_{t-2} \in \mathcal{F}_{t-2}$. It follows immediately that $K_{t-2}$ must be constant and from **A3**, that

$$\boldsymbol{\alpha}_0^*(\varphi_m) = \boldsymbol{\alpha}_\theta^*(\varphi_m), \quad \forall m \ge 1.$$

By **A8** (b) we deduce that $\boldsymbol{\alpha}_\theta = \boldsymbol{\alpha}_0$. From Eq. (A.16) we have, moreover, that

$$\langle \delta_0, \varphi_m\rangle + \langle \boldsymbol{\beta}_0(\sigma_{t-1}^2(\theta)), \varphi_m\rangle = \langle \delta_\theta, \varphi_m\rangle + \langle \boldsymbol{\beta}_\theta(\sigma_{t-1}^2), \varphi_m\rangle \quad \text{a.s.}$$

or, equivalently, that

$$\langle \delta_0, \varphi_m\rangle + \langle \sigma_{t-1}^2, \boldsymbol{\beta}_0^*(\varphi_m)\rangle = \langle \delta_\theta, \varphi_m\rangle + \langle \sigma_{t-1}^2(\theta), \boldsymbol{\beta}_\theta^*(\varphi_m)\rangle \quad \text{a.s.}$$

It follows that

$$\langle \delta_0, \varphi_m\rangle + \langle \delta_{\theta_0} + \boldsymbol{\alpha}_0(y_{t-2}^2) + \boldsymbol{\beta}_{\theta_0}(\sigma_{t-2}^2), \boldsymbol{\beta}_0^*(\varphi_m)\rangle \\ = \langle \delta_\theta, \varphi_m\rangle + \langle \delta_\theta + \boldsymbol{\alpha}_\theta(y_{t-2}^2) + \boldsymbol{\beta}_\theta(\sigma_{t-2}^2(\theta)), \boldsymbol{\beta}_\theta^*(\varphi_m)\rangle \quad \text{a.s.}$$

Because $\boldsymbol{\alpha}_\theta = \boldsymbol{\alpha}_0$ we deduce, with obvious notation, that

$$\langle \boldsymbol{\alpha}_0(y_{t-2}^2), [\boldsymbol{\beta}_0^*(\varphi_m) - \boldsymbol{\beta}_\theta^*(\varphi_m)]\rangle = K_{t-3}, \quad \text{a.s.}$$

or, equivalently, that

$$\langle \eta_{t-2}^2, \sigma_{t-2}^2(\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}_0^*)(\varphi_m)\rangle = \langle \eta_{t-2}^2, \sigma_{t-2}^2(\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}_\theta^*)(\varphi_m)\rangle + K_{t-3}, \quad \text{a.s.}$$

With similar arguments as in the above we, thus, obtain $(\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}_0^*)(\varphi_m) = (\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta})(\varphi_m)$, for all $m \geq 1$. By **A8** (c), this entails $\boldsymbol{\beta}_\theta = \boldsymbol{\beta}_0$. Finally, Eq. (A.16) reduces to $\langle \delta_0, \varphi_m \rangle = \langle \delta_\theta, \varphi_m \rangle$, which, by **A8** (a), implies that $\delta = \delta_0$. We can conclude that $\theta = \theta_0$.

Recall that $\widehat{\theta}_n$ denotes the minimiser of $\widetilde{Q}_n$ over the whole space $\Theta$ which is not computable in practise due to our assumption that $\Theta$ is infinite dimensional. Let us denote $\|x\|$ the $l^2$-norm of some square summable sequence $x = (x_1, x_2, \dots)$. We have that

$$\|\widehat{\theta}_n^{N_n} - \theta_0\| \leq \|\widehat{\theta}_n^{N_n} - \widehat{\theta}_n|_{\Theta_{N_n}}\| + \|\widehat{\theta}_n|_{\Theta_{N_n}} - \widehat{\theta}_n\| + \|\widehat{\theta}_n - \theta_0\|, \tag{A.17}$$

where $|_{\Theta_N}$ denotes the projection $x \mapsto (x_1, \dots, x_N)$. Up to now, we showed that the third term of (A.17) converges almost surely to zero. The second term is equal to $\left(\sum_{j>N_n} \widehat{\theta}_{n,j}^2\right)^{1/2}$ where $\widehat{\theta}_{n,j}$ simply denotes the $j$th term of the sequence $\widehat{\theta}_n$ which is supposed to be in $l^2$. We can further bound the square of this quantity by $\sup_{\ell \geq 1} \sum_{j>N_n} \widehat{\theta}_{\ell,j}^2$. To show that this term converges to zero we apply the tightness Lemma 14 in Cerovecki and Hörmann (2017) with $p_j^{(n)} = \widehat{\theta}_{n,j}^2$ and $p_j^{(0)} = \theta_{0,j}^2$. Finally, from the compactness of $\Theta$ we know that there exists a subsequence $(\widehat{\theta}_{n_\ell}^{N_{n_\ell}})_{\ell \geq 1}$ that converges in $\ell^2$ to $x$, say, and observe that by definition

$$\widetilde{Q}_{n_\ell}(\widehat{\theta}_{n_\ell}^{N_{n_\ell}}) \leq \widetilde{Q}_{n_\ell}(\widehat{\theta}_{n_\ell}|_{\Theta_{N_{n_\ell}}}), \quad \text{for all } \ell \geq 1.$$

Now, we have already shown that $\widehat{\theta}_{n_\ell}|_{\Theta_{N_{n_\ell}}} \to \theta_0$ a.s. when $\ell \to \infty$. Since $\widetilde{Q}_n(\theta) \to Q(\theta)$ a.s. and uniformly in $\theta$, we obtain that $Q(x) \leq Q(\theta_0)$, a.s. This shows that $x = \theta_0$ and, thus, that the first term on the right-hand side of (A.17) converges a.s. to zero. □

## References


Aue, A., Dubart Norinho, D., Hörmann, S., 2015. On the prediction of stationary functional time series. J. Amer. Statist. Assoc. 110, 378–392.
Aue, A., Hörmann, S., Horváth, L., Reimherr, M., 2009. Break detection in the covariance structure of multivariate time series models. Ann. Stat. 37, 4046–4087.
Aue, A., Horváth, L., Pellatt, D.F., 2017. Functional generalized autoregressive conditional heteroskedasticity. J. Time Series Anal. 38, 3–21.
Aue, A., Klepsch, J., 2017. Estimating invertible functional time series. Funct. Stat. Related Fields 51–58.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31, 307–327.
Bollerslev, T., 1990. Modellling the coherence in short–run nominal exchangerates: a multivariate generalized ARCH model. Rev. Econom. Stat. 72, 498–505.
Boubacar Maïnassara, Y., Saussereau, B., 2018. Diagnostic checking in multivariate ARMA models with dependent errors using normalized residual autocorrelations. J. Amer. Statist. Assoc. 113, 1813–1827.
Cardot, H., Mas, A., Sarda, P., 2007. CLT in functional linear regression models. Probab. Theory Related Fields 138, 325–36.1.
Cerovecki, C., Hörmann, S., 2017. On the CLT for discrete Fourier transforms of functional time series. J. Multivariate Anal. 154, 282–295.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econ. 7, 174–196.
Eichler, M., van Delft, A., 2017. Locally stationary functional time series. arXiv preprint 1602.05125v2.
Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50, 987–1008.
Escanciano, J.C., 2009. Quasi-maximum likelihood estimation of semi-strong GARCH models. Econometric Theory 25, 561–570.
Francq, C., Zakoïan, J.M., 2011. GARCH Models: Structure, Statistical Inference and Financial Applications. John Wiley & Sons.
Francq, C., Zakoïan, J.M., 2012. QML estimation of a class of multivariate asymmetric GARCH models. Econometric Theory 28, 179–206.
Hörmann, S., Horváth, L., Reeder, R., 2013. A functional version of the ARCH model. Econometric Theory 29, 267–288.
Hörmann, S., Kokoszka, P., 2010. Weakly dependent functional data. Ann. Statist. 38, 1845–1884.
Horváth, L., Kokoszka, P., 2014. Inference for Functional Data with Applications, Vol. 200. Springer Science & Business Media.
Horváth, L., Kokoszka, P., Rice, G., 2014. Testing stationarity of functional time series. J. Econometrics 179, 66–82.
Kingman, J.F.C., 1973. Subadditive ergodic theory. Ann. Probab. 1, 883–899.
Klepsch, J., Klüppelberg, C., 2017. An innovations algorithm for the prediction of functional linear processes. J. Multivariate Anal. 155, 252–271.
Paparoditis, E., 2017. Sieve bootstrap for functional time series. arXiv preprint 1609.06029v2.
Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. J. Econometrics 160, 246–256.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, second ed. Springer, New York.
Zhu, T., Politis, D.N., 2017. Kernel estimation of first-order nonparametric functional autoregression model and its bootstrap approximation. Electron. J. Stat. 11, 2876–2906.