Functional GARCH models: the quasi-likelihood approach and its applications

Cerovecki, Clément and Francq, Christian and Hormann, Siegfried and Zakoian, Jean-Michel

Université libre de Bruxelles, CREST and University of Lille, Université libre de Bruxelles, CREST and University of Lille

18 January 2018

Functional GARCH models: the quasi-likelihood approach and its applications

Clément Cerovecki[*], Christian Francq[†], Siegfried Hörmann[‡] and Jean-Michel Zakoïan[§]

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.

JEL Classification: C13, C32 and C58.
Keywords: Functional time series, High-frequency volatility models, Intraday returns, Functional QMLE, Stationarity of functional GARCH.

[*]Université libre de Bruxelles, Boulevard du Triomphe, CP210 B-1050 Brussels - Belgium Office: NO O.9.211. *E-Mail address*: clement.cerovecki@ulb.ac.be

[†]CREST and University of Lille, BP 60149, 59653 Villeneuve d'Ascq cedex, France. *E-Mail address*: christian.francq@univ-lille3.fr

[‡]Université libre de Bruxelles, Boulevard du Triomphe, CP210 B-1050 Brussels - Belgium Office: NO O.9.115

[§]Corresponding author: Jean-Michel Zakoïan, University of Lille and CREST, 5 Avenue Henry Le Chatelier, 91120 Palaiseau, France. *E-Mail address*: zakoian@ensae.fr, Phone number: 33.1.70.26.68.46.

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 Zakoian (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 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 $(X_t \colon 1 \leq t \leq n)$, where each random object X_t is a curve $(X_t(u) \colon u \in [0,1])$. 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 τ-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. (2016) 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. (2016) 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 know from scalar GARCH theory that the least-squares estimators lack efficiency.

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 instrumental functions. We give regularity conditions for consistency and asymptotic normality. As a result that is aside from this study, we obtain the consistency and asymptotic normality of a semi-strong (i.e. with non-iid innovations) multivariate CCC-GARCH. We also obtain a sufficient condition for existence of stationary functional GARCH processes (Section 2.2) which generalises Aue et al. (2016). 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 risk 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 most of the 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 $\boldsymbol{\alpha\beta} := \boldsymbol{\alpha} \circ \boldsymbol{\beta}$ 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 $\|\boldsymbol{\alpha}\| := \sup_{\|x\| \leq 1} \|\boldsymbol{\alpha}(x)\|$, is a Banach space. This norm is sub-multiplicative, i.e. $\|\boldsymbol{\alpha\beta}\| \leq \|\boldsymbol{\alpha}\| \, \|\boldsymbol{\beta}\|$. In some places we also make use of the supremum norm: $\|x\|_\infty = \inf\{a > 0, \ |x(u)| < a$, for λ–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 $H^+ = \{x \in H : x(u) \geq 0$, for almost every $u \in [0,1]\}$ and $H^+_* = \{x \in H : x(u) > 0$, 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 $k \geq 2$, 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 σ.

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 $\mathcal{F}_t$ generated by $(\eta_s)_{s\leq t}$. 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 δ. 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 ϕ 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 δ.

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 (η_t) such that $\mathrm{Var}(\eta_t(u)) = 1$ for all $u \in [0, 1]$. We can then interpret the region

$\{[-2\sigma_t(u), 2\sigma_t(u)] \colon u \in [0,1]\}$ as the prediction interval of $y_t(u)$ at (approximate) level $\alpha = .05$. We show these curves and their estimation in Figure 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:

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

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

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

Here, the operator $\mathbf{\Upsilon}_t$ is the pointwise multiplication by η_t^2, i.e. $H \ni x \longmapsto x\eta_t^2$. All 0's in the definition of the matrix $\mathbf{\Psi}_t$ 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 $\mathbf{\Upsilon}_t$ is almost surely bounded. Indeed, we have that

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

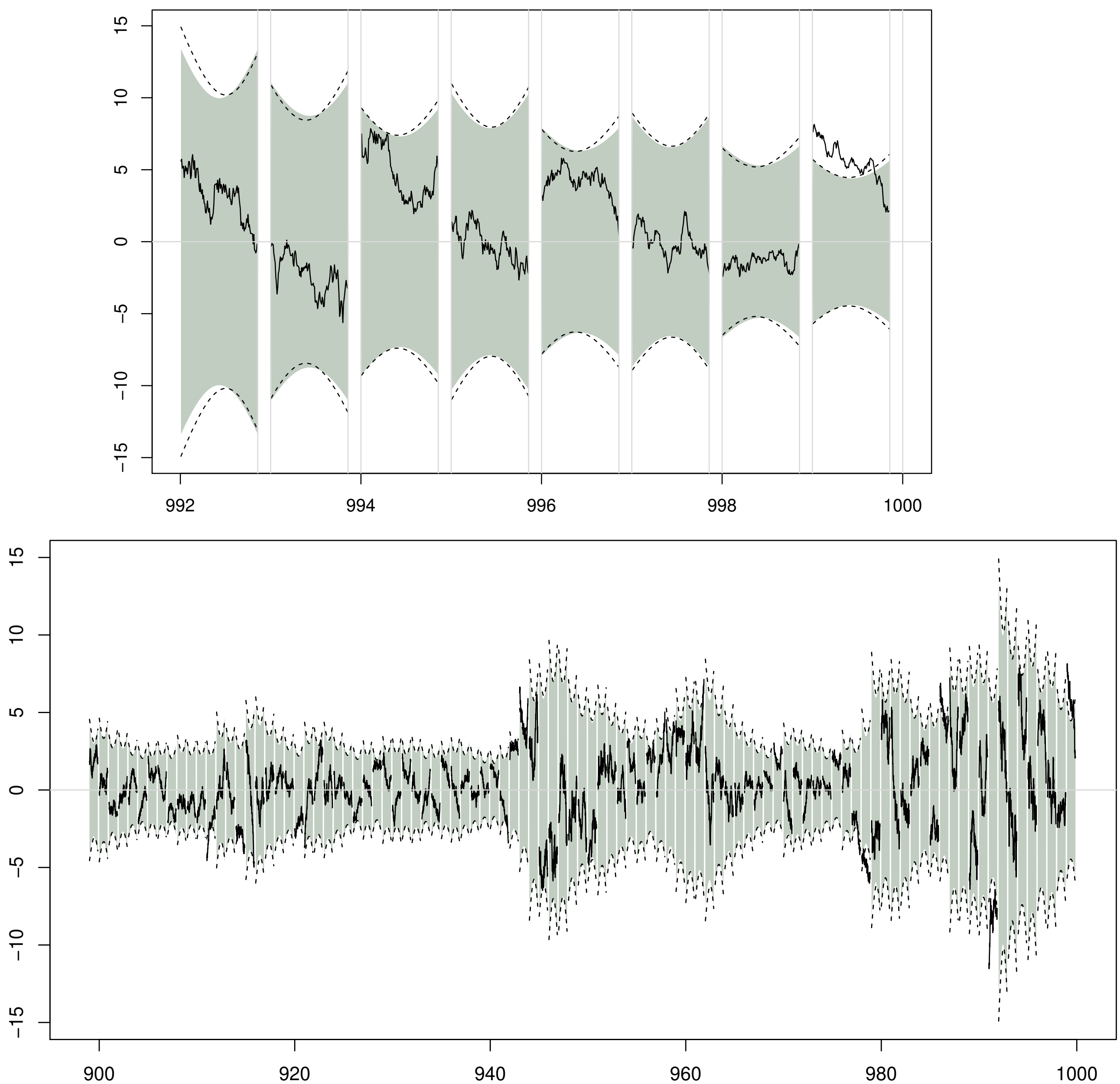

Figure 1: Solid lines represent the simulated process y_t, the shaded area is the region $\{[-2\sigma_t(u), 2\sigma_t(u)]\colon u \in [0,1]\}$. The dashed lines are estimators $\pm 2\widetilde{\sigma}_t(\hat{\theta})(u)$.

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, almost surely,

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

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

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* $\gamma < 0$.

Proof. By iterating (2.4), we formally get that

$$\underline{z}_t = \underline{b}_t + \sum_{k=1}^{\infty}\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_{t-k+1}(\underline{b}_{t-k}). \tag{2.9}$$

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

$$\limsup_{t\to\infty}\frac{1}{t}\log\|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_{t-k+1}(\underline{b}_{t-k})\| \leq \gamma + \limsup_{t\to\infty}\frac{1}{t}\log\|\underline{b}_{t-k}\|, \quad \text{a.s.} \tag{2.10}$$

Since $E\log^+\|\underline{b}_{t-k}\| < \infty$ by (2.5), and $\|\underline{b}_{t-k}\| \geq \|\delta\| > 0$ the second summand 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 $\underline{\tilde{z}}_t^*$ 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

$$\|\underline{z}_t^* - \underline{z}_t\| \leq \|\underline{z}_t^N - \underline{z}_t\| + \|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_{t-N}\|\cdot\|\underline{z}_{t-N-1}^*\|. \tag{2.11}$$

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 (2.11) tends to zero in probability. Therefore, $P(\underline{z}_t^* = \underline{z}_t) = 1$. □

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 the series in (2.9)) will imply $\|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_{t-k+1}\|_* \to 0$. In the infinite-dimensional setup, however, norms are not equivalent, and choosing a different norm will also give a different value for the exponent γ. In a second step, one uses contraction properties of random matrices in order to conclude. To extend such results to linear operators is beyond the scope of this paper.

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

Proposition 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*

$$E\log\left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{t-1}+\boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1+\boldsymbol{\beta})\right\| < 0, \quad \textit{for some } t \geq 1.$$

Proof. 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. □

In their recent paper, Aue et al. (2016) obtained the condition

$$E\log\|\boldsymbol{\alpha}\boldsymbol{\Upsilon}_0+\boldsymbol{\beta}\|_{\mathcal{S}} < 0 \tag{2.12}$$

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,

$$\gamma \leq \frac{1}{t}E\log\left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_t+\boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1+\boldsymbol{\beta})\right\| \leq E\log\left\|\boldsymbol{\alpha}\boldsymbol{\Upsilon}_0+\boldsymbol{\beta}\right\| \leq E\log\|\boldsymbol{\alpha}\boldsymbol{\Upsilon}_0+\boldsymbol{\beta}\|_{\mathcal{S}}.$$

This shows that our condition is milder than that of Aue et al. (2016).

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

Proof. 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

$$E\underline{z}_t = E\underline{b}_t + \sum_{k=1}^{\infty} (E\boldsymbol{\Psi}_0)^k \, E\underline{b}_0, \tag{2.13}$$

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 $u \in [0,1]$. □

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

Proposition 3. *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 $E\|\sigma_t^2\|^s < \infty$.*

Proof. 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

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

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 Zakoian (2011), we then deduce that there exists an $0 < s < \tau$ such that $\varsigma := E\|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_1\|^s < 1$. From (2.9) we get that

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

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. □

3 Estimation

A difficulty in estimating FDA models is that the concept of likelihood does not exist. For this reason, QML estimation cannot be straightforwardly 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. As regards the latter, the functional QMLE will be shown to be consistent in a semi-parametric framework in which the distribution of η_t does not need to be specified.

3.1 Parametrisation

From observations $(y_t)_{1\le t\le n}$ of curves satisfying Model (2.1)–(2.2), we consider inference on the parameters δ, $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_q$ and $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_p$. 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 η_t.

We begin by assuming a specific parametrisation for Model (2.1)–(2.2). Let $\varphi_1,\dots,\varphi_M$ be linearly independent functions in H^+. We assume that there exists a non-negative valued vector $d=(d_1,\dots,d_M)'$ in $\mathbb{R}^M$, and non-negative valued 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 $i=1,\dots q$, $j=1,\dots p$. Note that $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ belong to $\mathcal{K}_H^+$.[1] We 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 θ_0. By convention, we index by zero all quantities evaluated at θ_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 $i=1,\dots q$, $j=1,\dots p$. To avoid confusion with the *parameter* θ we refer to δ, $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_q,\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_p$ as the *functional* parameters of the model. We assume that θ_0 belongs to a compact subset Θ of $\mathbb{R}_+^{M+(p+q)M^2}$.

Remark 2. The implication of (3.2) is that the volatility process $(\sigma_t^2)_{t\in\mathbb{Z}}$ belongs to the M-dimensional subspace of H spanned by $\varphi_1,\dots,\varphi_K$. This is also assumed in Aue et al. (2016). 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

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

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

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 initial values $y_0, \dots, y_{-q+1}$ and $\tilde{\sigma}_0, \dots, \tilde{\sigma}_{-p+1}$ in H. Note that the positivity of the baseline functions φ_k ensures that the scalar products in (3.5) are positive and, thus, $\tilde{Q}_n(\theta)$ is well defined and reaches its minimum on the compact set Θ. 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 process $(h_t(\theta))_{t\in\mathbb{Z}}$ as the stationary and ergodic solution of the following equation

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

where $Y_t^{<2>} = (\langle y_t^2, \varphi_1\rangle, \dots \langle y_t^2, \varphi_M\rangle)'$, $\mathfrak{d} = \Phi d$ 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 $\theta_0 \in \Theta$, Θ 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 κ,

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

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 $\inf_{\theta\in\Theta} \mathfrak{d} > 0$ componentwise, and for all $\theta \in \Theta$, the matrix $\mathfrak{B}(z) = I_M - \sum_{j=1}^{p} \Phi B_i z^i$ is invertible for $|z| \leq 1$.

Theorem 2. *Under* (3.1)–(3.2) *and Assumptions* **A1-A5**, *the QMLE of θ_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 $E\|\eta_0\|_\infty^4 < \infty$.

Letting $\mathcal{N}_K(\mu, \Sigma)$ denote a K-variate normal random vector with mean μ and covariance matrix Σ, 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*

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

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

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

Using this, it is not difficult to obtain estimates for I and J.

3.3 Choice of instrumental functions

The instrumental functions $\varphi_1, \ldots, \varphi_M$ must satisfy the 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 (2006).

In order to avoid using a specific set of instrumental functions, we propose a heuristic data-driven method. 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. 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 (2012) 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 δ, $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ are spanned by the finite set of functions $\mu, \psi_1, \ldots, \psi_{M-2}$. However, these functions are not non-negative. We, thus, 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:

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

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.

4 Extension to infinite-dimensional parameter space

Assuming a finite-dimensional parametrisation (3.2) may appear to be not entirely satisfactory 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 δ, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be parametrised by some infinite-dimensional parameter $\theta \in \Theta$, in other words we let $M = \infty$ in (3.2). This parameter space is assumed to be a compact subset of l^2 (the set of square summable sequences).

Our new estimator is defined as

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

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) and we set

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

where $(w_m)_{m\geq 1}$ is a non-negative 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 Identifiability. For all $m \geq 1$ and $\theta \in \Theta$ we have that

(a) $\langle \delta_0, \varphi_m \rangle = \langle \delta, \varphi_m \rangle$ implies $\delta = \delta_0$.

(b) $\boldsymbol{\alpha}_0^*(\varphi_m) = \boldsymbol{\alpha}^*(\varphi_m)$ implies $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$.

(c) $(\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}_0^*)(\varphi_m) = (\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}^*)(\varphi_m)$ implies $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

A9 There exists a 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 δ_θ is uniformly bounded from below by some constant $c > 0$.

Remark 3. Note that, if $M < \infty$, $p = q = 1$, and $\mathfrak{A}_0$ is invertible then (3.2) implies **A8**.

Proposition 4. *Under* (3.1) *and assumptions* **A1**$-$**A3***, and* **A8** $-$**A9***, the QMLE of* θ_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 recursion 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 of it. This model provides a very simple interpretation of the function δ. Indeed, we can show that the curve of δ 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 η_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 4 can be applied to model (4.2). Indeed, we can easily choose a compact subset Θ of ℓ^2 and an innovation process $(\eta_t)_{t\in\mathbb{Z}}$ such that Assumptions **A1**$-$**A3** are satisfied. Let $(\varphi_m)_{m\geq 1}$ be any family of non-negative functions spanning H. If we further 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 $\delta(u) = \exp\left(\sum_{k=1}^\infty d_k\psi_k(u)\right) \geq \exp\left(-\kappa \sum_{k=1}^\infty |d_k|\right) > 0$. Now, since $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, it easy to see that **A8** (b) is satisfied as well as **A8** (c), provided that $a_0 \neq 0$. Assumption **A8** (a) follows from the fact that $(\varphi_m)_{m\geq 1}$ is a basis of the space H. Then, for any weight sequence $(w_m)_{m\geq 1}$ which satisfies **A9** we get by Proposition 4 that the QMLE of model (4.2) is strongly consistent.

5 Empirical results

5.1 Simulations

We will first compare the Least Squares Estimator (`LSE`) of Aue et al. (2016) and our `QMLE` that is defined in (3.4). Next, we compare the QML with given instrumental functions φ_m to the data-driven procedure described in Section 3.3 (we then refer to `QMLE*`).

	d	a	b
sd	2.4e-05	0.01	0.031
bias	1.6e-05	0.0041	0.011

Table 1: Performance of `LSE`.

	d	a	b
sd	7.7e-06	0.0051	0.0093
bias	3.1e-06	0.0017	0.0036

Table 2: Performance of `QMLE`.

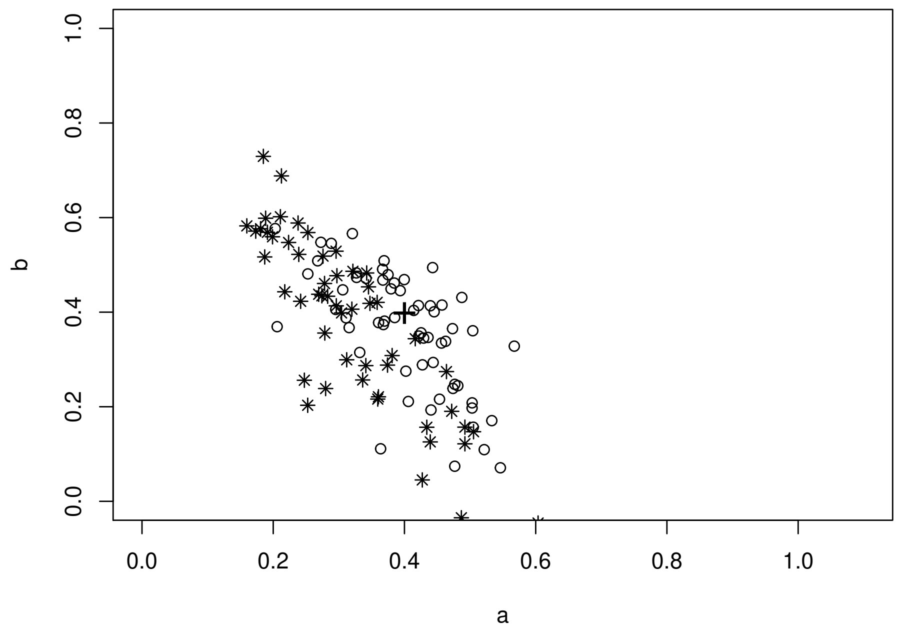

Figure 2: Estimates of a and b, with $*$ `LSE` and $\circ$ `QMLE`, and $+$ indicates the true values.

5.1.1 Example 1

The first setup is taken from Aue et al. (2016). They consider a GARCH(1,1) model with

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

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. (2016) project on one basis function $\varphi_1(u) = \sqrt{30}u(1-u)$, $u \in [0,1]$. It follows that $K_{\boldsymbol{\alpha}}(u,v) = a\,\varphi_1(u)\varphi_1(v)$, with $a = 0.4$ and $K_{\boldsymbol{\beta}}(u,v) = b\,\varphi_1(u)\varphi_1(v)$, with $b = 0.4$. Note that δ is not spanned by $\varphi_1(u)$ and that $d = \langle\delta, \varphi_1\rangle \approx .009$. It is assumed that φ_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 100 Monte-Carlo replications of our estimation experiment. The results of our simulations are displayed in Tables 1–2 and in Figure 2. We see that standard deviation and bias differ by a factor of 2 to 3 in favour of the `QMLE` methods.

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 $\delta(u) = (u - .5)^2 + .1$,

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

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 are discarded.

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

1. `QMLE`: Bernstein polynomials, which are a special case 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.

We fix $M = 4$, then the subspace spanned by the Bernstein polynomials (of order 3) 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 to avoid an explosive solution, $0 \leq b_{k\ell} \leq (M \cdot \max_{1 \leq m \leq M} \|\varphi_m\|)^{-1}$, for all $k, \ell = 1, \ldots, M$.

We have represented the functional parameters δ and $\boldsymbol{\alpha}$ (its kernel $K_{\boldsymbol{\alpha}}$) in Figure 3 together with their `QMLE`. In order to compare the performance of the two procedures, we ran $N = 100$ Monte-Carlo replications of our estimation experiment with sample size $n = 1000$. The results of our simulations are displayed in Table 3. We show the relative mean squared deviations, defined as

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

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

	δ	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$
`QMLE`	0.45	0.46	0.55
`QMLE*`	0.51	0.33	0.44

Table 3: Relative mean squared deviations for the corresponding functional parameters.

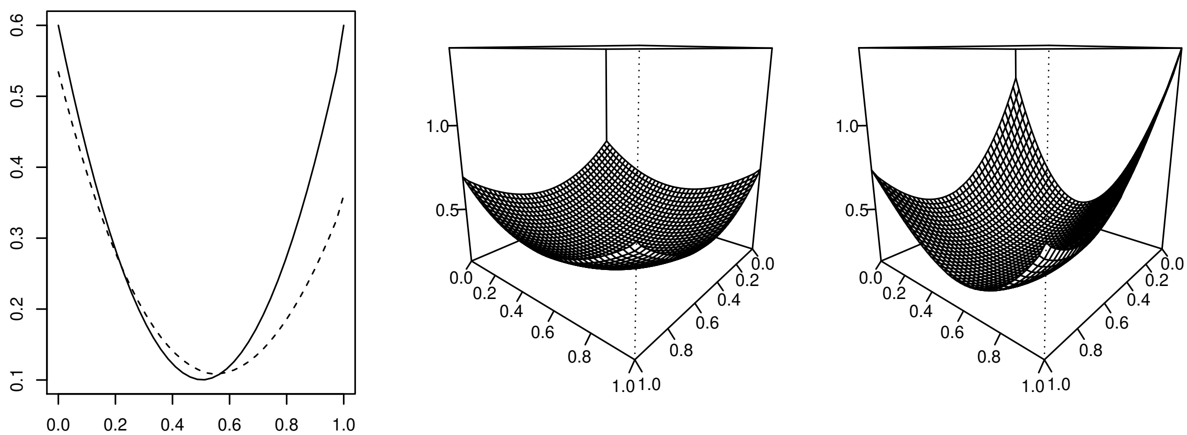

Figure 3: From left to right: the intercept function δ (solid line) compared to its estimation $\hat{\delta}$ (dashed line), the theoretical kernel $K_{\boldsymbol{\alpha}}(u, v)$ and the estimated kernel $K_{\widehat{\boldsymbol{\alpha}}}(u, v)$.

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 Figure 4. The functional GARCH

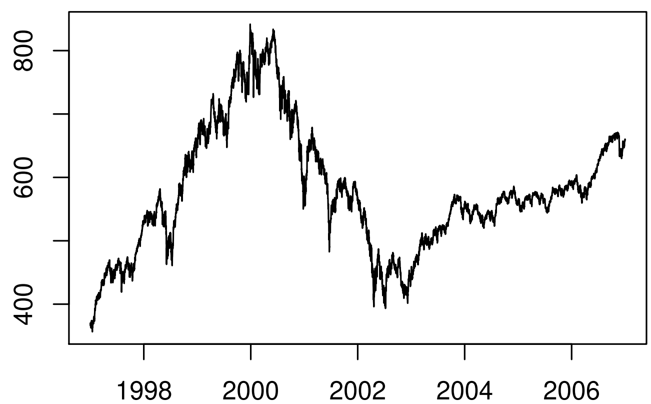

Figure 4: Raw data for S&P 100 index between 1997 and 2007.

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 τ-minute returns $y_t(u)$ and the intraday returns $\tilde{y}_t(u)$ defined by

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

For $y_t(u)$ we used $\tau = 20\,\text{min}$[2]. For the instrumental functions we used, as in Section 5.1.2 the Bernstein polynomials with $M = 4$. We computed the `LSE` estimator to get an initial value of the parameter in the optimisation routine. The resulting empirical volatility curves are displayed in Figures 5 and 6 for y_t and in Figures 7 and 8 for $\tilde{y}_t$. In light of (2.3) and the related discussion, we plotted the curves $\tilde{\sigma}_t(\widehat{\theta}_n)(u) \cdot \widehat{Q}^{\hat{\eta}(u)}_{1-\alpha/2}$. 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 can observe the sensitivity to shocks of the volatility process and its persistence. The persistence seems stronger in Figure 8 than in Figure 6, whereas the rise of the volatility after a shock is more evident Figure 5 than in Figure 6. This is in line with a large value of $\|\widehat{\boldsymbol{\beta}}\|$ for $\tilde{y}_t$ (see Table 4).

[2] Note that the functional approach is not very appropriate when τ is less than, say 10 min, since then, the curves become very irregular and, thus, there is no chance to capture the intraday dynamic by smoothing the signal.

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

Table 4: Norms of the estimated parameters.

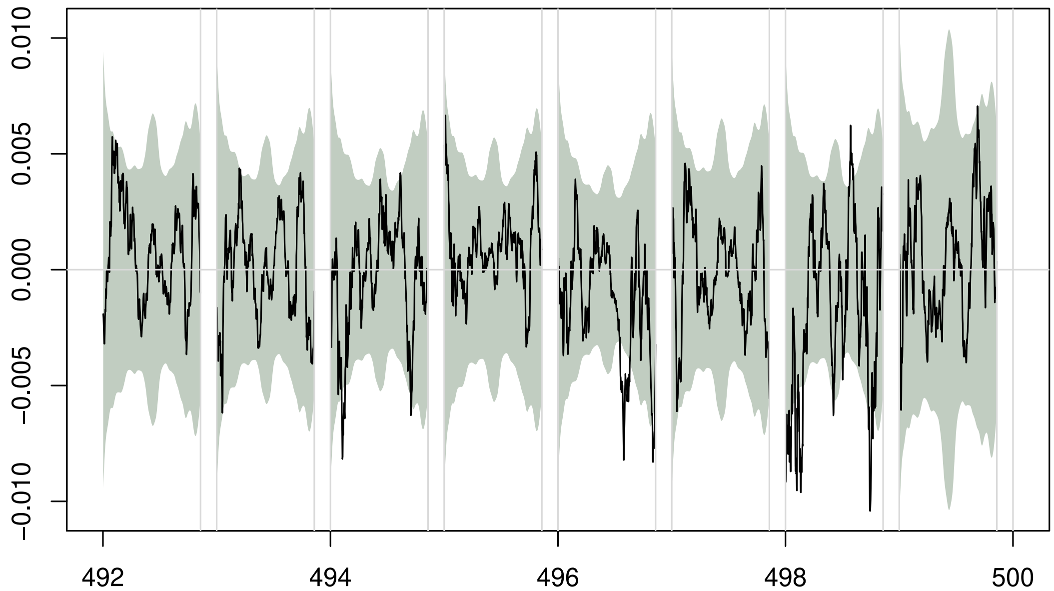

Figure 5: Predicted volatility (shaded area) for y_t (8 days).

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 τ as in the definition of y_t in (5.3) we remark that

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

At time $t-1$, the optimal predictor of 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

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

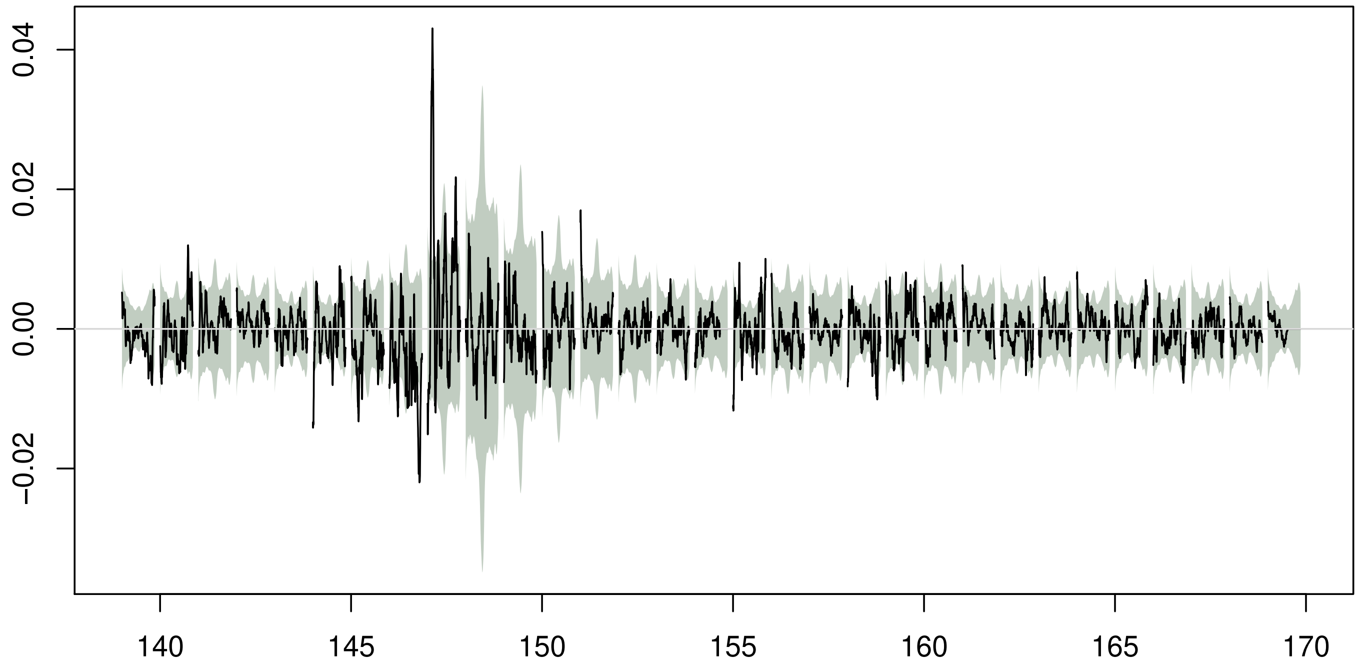

Figure 6: Predicted volatility (shaded area) for y_t (31 days).

where $\widehat{\theta}$ is the `QMLE` computed with the sub-sample $y_1, \ldots, y_{t-1}$. In Figure 9, we have plotted 41 one-day ahead predictions of $\widetilde{\mathrm{RV}}_t$ against RV_t.

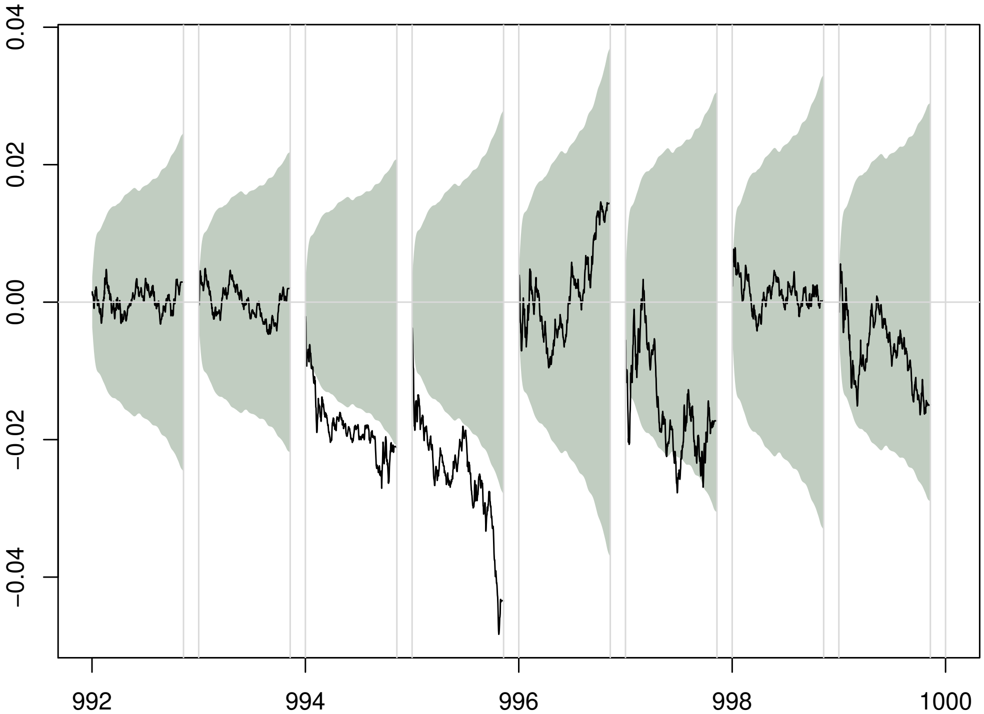

Figure 7: Predicted volatility (shaded area) for $\tilde{y}_t$ (8 days).

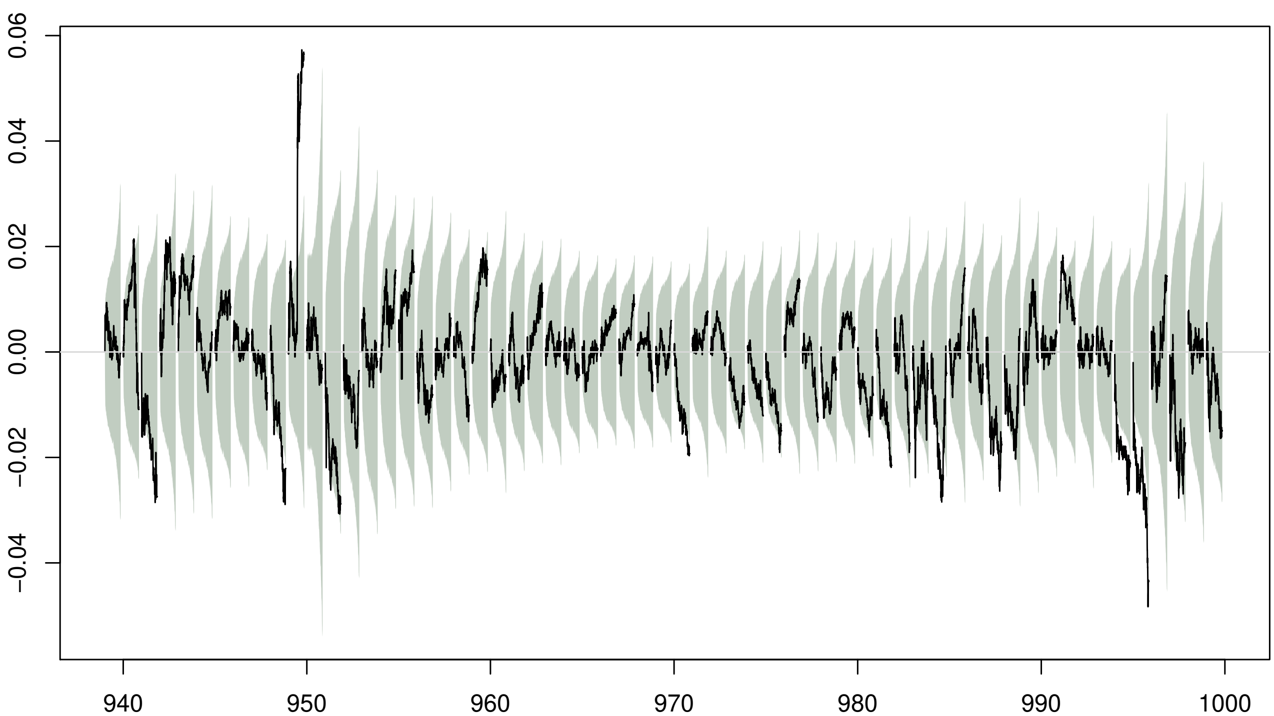

Figure 8: Predicted volatility (shaded area) for $\tilde{y}_t$ (61 days).

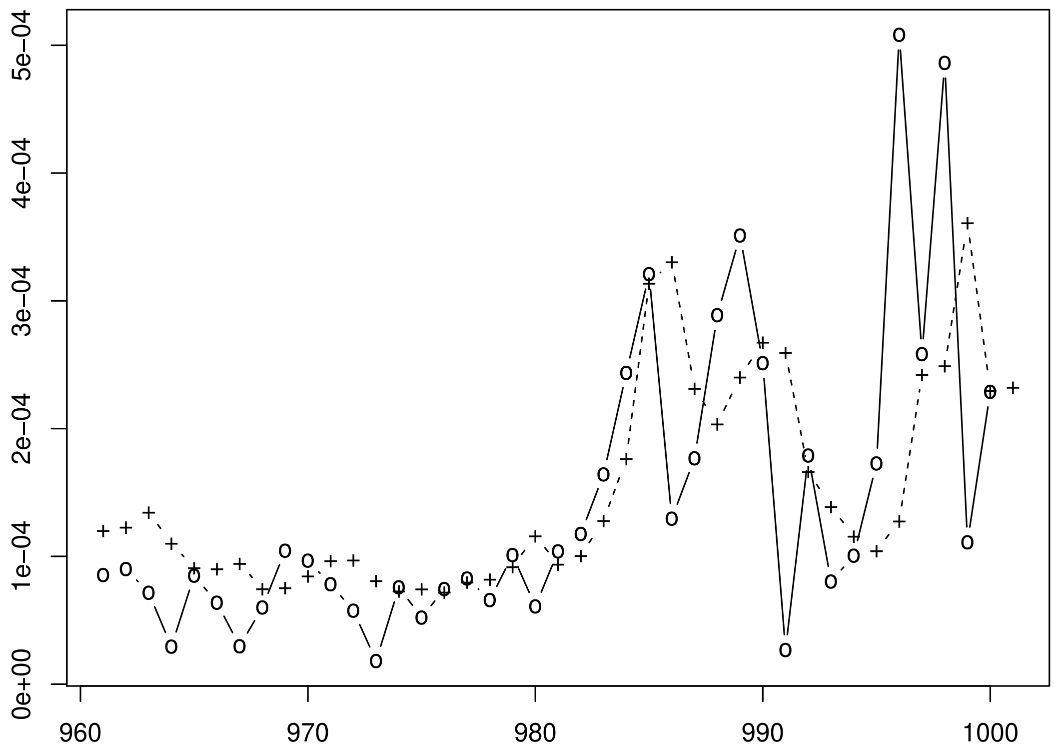

Figure 9: Predicted (+) and true realised volatility (o).

Appendix

We start to show asymptotic results for CCC-GARCH models which will be used to prove Theorems 2 and 3. Similar results exist in the literature but with independent innovations. 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.

5.3 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:

$$\epsilon_t = H_t^{1/2}\nu_t, \tag{5.6}$$

$$H_t = D_t R D_t \quad \text{with} \quad D_t = \Big(\mathrm{diag}(h_t)\Big)^{1/2}, \tag{5.7}$$

$$h_t = \mathfrak{d} + \sum_{i=1}^{q} \mathfrak{A}_i \epsilon_{t-i}^{[2]} + \sum_{j=1}^{p} \mathfrak{B}_j h_{t-j}, \tag{5.8}$$

where $\epsilon_t^{[2]} = (\epsilon_{t,1}^2,\ldots,\epsilon_{t,M}^2)$. Here R is an $M\times M$ correlation matrix, $\mathfrak{A}_i$ and $\mathfrak{B}_j$ are $M\times M$ matrices with positive elements, the components of the M-vector $\mathfrak{d}$ are strictly positive. 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 ξ_0 is defined by

$$\hat{\xi}_n = \arg\min_{\xi\in\Xi} \frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_t(\xi), \quad \tilde{\ell}_t(\xi) = \epsilon_t'\tilde{H}_t^{-1}\epsilon_t + \log|\det(\tilde{H}_t)|, \tag{5.9}$$

where $\tilde{H}_t$ is defined recursively using (5.8) and some initial values.

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 ξ belonging to a compact parameter set Ξ. 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 Ξ is a compact set.

A*2 (ϵ_t) is a strictly stationary and ergodic solution of Model (5.6)-(5.8), 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.

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

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

$\mathbf{A}^*\mathbf{7}$ $E\|\nu_t\nu_t'\|^{2(1+w)} < \infty$, 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* $\mathbf{A}^*\mathbf{0}$–$\mathbf{A}^*\mathbf{5}$ *the QMLE of* ξ_0 *as defined in* (5.9) *is strongly consistent, i.e.* $\widehat{\xi}_n \to \xi_0$ a.s.

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

$$\sqrt{n}(\hat{\xi}_n - \xi_0) \xrightarrow{d} \mathcal{N}_{M+(p+q)M^2+M(M-1)/2}\left(0, J^{-1}IJ^{-1}\right) \tag{5.10}$$

where $I = \mathrm{Var}\left(\frac{\partial \ell_t(\xi_0)}{\partial \xi}\right)$ *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 Zakoian (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 $\mathbf{A}^*\mathbf{0}$. 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 $\mathbf{A}^*\mathbf{3}$ we may conclude. □

Proof of Theorem 5. As one can see in the proof of Theorem 11.8 in Francq and Zakoian (2011), the independence of the innovations is only used to show the existence of moments of the theoretical criterion ℓ_t, or of its derivative at ξ_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 ν_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 $\mathbf{A}^*\mathbf{3}$, the moment condition on ν_t has to be strengthened as in $\mathbf{A}^*\mathbf{7}$. More precisely, for $i, j \leq s$, we use Hölder's inequality to get that

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

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. □

5.4 Proofs of the results of Section 3

In view of representation (3.7), we build a sequence (ϵ_t) satisfying the CCC-GARCH model of the previous section. Let $(r_t)_{t\in\mathbb{Z}}$ be an i.i.d. sequence of M-dimensional vectors, whose components are independent Rademacher variables. Let $\epsilon_t = \{\text{diag}(Y_t^{<2>})\}^{1/2} r_t$ and let $\mathcal{F}_t$ the σ-field generated by $(r_t, \{\eta_{t-u}, u \geq 0\})$. Let $D_t = \{\text{diag}(\langle \sigma_t^2, \varphi_m \rangle)\}^{1/2}$ and let $\nu_t = D_t^{-1}\epsilon_t$. Note that $\epsilon_t^{[2]} = (\epsilon_{t,1}^2, \dots, \epsilon_{t,M}^2) = Y_t^{<2>}$. It follows that (5.6)-(5.8) hold with $R = I_M$. Let

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

where $\otimes$ denotes the usual Kronecker product of matrices. Since Φ is non-singular (this follows from the linear independence of the functions $\varphi_1, \dots, \varphi_M$), the transformation T which maps θ to ξ is bijective. By choosing $\Xi = T(\Theta)$, the QML estimator $\hat{\xi}_n$ defined in (5.9) satisfies $\hat{\xi}_n = T(\hat{\theta}_n)$. Clearly, Theorems 4-5 can be straightforwardly adapted when $R = I$ 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 $\epsilon_t = \{\text{diag}(Y_t^{<2>})\}^{1/2} r_t$ satisfies assumptions $\mathbf{A^*0}$–$\mathbf{A^*5}$.

Assumption $\mathbf{A^*0}$ follows from Proposition 3, noting that

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

where $\|.\|$ denotes the euclidean norm. Assumption $\mathbf{A^*1}$ is obviously satisfied. By construction $\epsilon_t \in \mathcal{F}_t$ and satisfies (5.6)-(5.8). The stationarity and ergodicity of ϵ_t readily follows from $\mathbf{A2}$. Thus $\mathbf{A^*2}$ holds. The first part of $\mathbf{A^*3}$ 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 $\mathbf{A^*3}$ 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

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

Assumption $\mathbf{A3}$ 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 $\sigma_t^2(u) \geq \delta_0(u) > 0$, for all $u \in [0,1]$ and the function $\varphi_1, \dots, \varphi_M$ are linearly independent, we can conclude that $x = 0$. The remaining assumptions $\mathbf{A^*4}$ and $\mathbf{A^*5}$ 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 $\mathbf{A6}$ implies that

A***6**. Next, we have

$$\begin{aligned}
E\|\nu_t\nu_t'\|^2 &\le \sum_{k,m=1}^{M} E\nu_{t,k}^2\nu_{t,m}^2 \\
&\le \sum_{k,m=1}^{M} \left(E\nu_{t,k}^4 E\nu_{t,m}^4\right)^{1/2} \\
&= \sum_{k,m=1}^{M} \left(E\Big[E[\langle\sigma_t^2\eta_t^2,\varphi_k\rangle^2|\mathcal{F}_{t-1}]/\langle\sigma_t^2,\varphi_k\rangle^2\Big]E\Big[E[\langle\sigma_t^2\eta_t^2,\varphi_m\rangle^2|\mathcal{F}_{t-1}]/\langle\sigma_t^2,\varphi_m\rangle^2\Big]\right)^{1/2} \\
&\le M^2 E\|\eta_t\|_\infty^4,
\end{aligned}$$

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

$$E\left\|\frac{\partial \ell_t(\theta_0)}{\partial\theta}\frac{\partial \ell_t(\theta_0)}{\partial\theta'}\right\| < \infty \quad \text{and} \quad E\left\|\frac{\partial^2 \ell_t(\theta_0)}{\partial\theta\partial\theta'}\right\| < \infty, \tag{5.11}$$

we compute

$$\frac{\partial \ell_t(\theta)}{\partial\theta} = \sum_{m=1}^{M}\left(1 - \frac{Y_{t,m}^{<2>}}{h_{t,m}}\right)\frac{1}{h_{t,m}}\frac{\partial h_{t,m}}{\partial\theta}, \tag{5.12}$$

$$\begin{aligned}
\frac{\partial^2 \ell_t(\theta)}{\partial\theta\partial\theta'} &= \sum_{m=1}^{M}\left(1 - \frac{Y_{t,m}^{<2>}}{h_{t,m}}\right)\frac{1}{h_{t,m}}\frac{\partial^2 h_{t,m}}{\partial\theta\partial\theta'} \\
&\quad + \sum_{m=1}^{M}\left(2\frac{Y_{t,m}^{<2>}}{h_{t,m}} - 1\right)\frac{1}{h_{t,m}^2}\frac{\partial h_{t,m}}{\partial\theta}\frac{\partial h_{t,m}}{\partial\theta'}.
\end{aligned} \tag{5.13}$$

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

$$\frac{Y_{t,m}^{<2>}}{h_{t,m}} = \frac{\langle y_t^2,\varphi_m\rangle}{\langle\sigma_t^2,\varphi_m\rangle} = \frac{\int \sigma_t^2(u)\eta_t^2(u)\varphi_m(u)du}{\int \sigma_t^2(u)\varphi_m(u)du} \le \sup_{u\in[0,1]} \eta_t^2(u) = \|\eta_t^2\|_\infty. \tag{5.14}$$

This last quantity is independent of $\mathcal{F}_{t-1}$, which readily implies that (5.11) reduces to prove that

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

for all $m = 1,\dots,M$ and $i = 1,\dots,M+(p+q)M^2$. This can be established with Proposition 3. □

Proof of Proposition 4. We first prove that $\widehat{\theta}_n \to \theta_0$ almost surely, where $\widehat{\theta}_n$ denotes the minimiser of $\widehat{Q}_n$ over the whole space Θ. Let $\ell_t(\theta)$ denote the theoretical criterion (involving the infinite past of y_t), defined as $\tilde{\ell}_t(\theta)$, but with $\tilde{\sigma}_t^2$ replaced by σ_t^2 given by the model recursion. Note that under (3.1) we have that

$$E\left[\langle y_t^2, \varphi_m\rangle \mid \mathcal{F}_{t-1}\right] = \langle \sigma_t^2, \varphi_m\rangle. \tag{5.15}$$

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.

(ii) $\langle \sigma_t^2, \varphi_m\rangle = \langle \sigma_t^2, \varphi_m\rangle$ 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 $\theta \in \Theta$, $\sigma_t^2 \geq \delta$. Hence, the non-negativity of the φ_m's and **A9** implies that there is a strictly positive constant c, such that $\langle \sigma_t^2, \varphi_m\rangle \geq c a_m$ for all $m \geq 1$. Furthermore, we have that $\|\sigma_t^2 - \tilde{\sigma}_t^2\| \leq K\rho^t$, where ρ is the supremum of $\|\boldsymbol{\beta}_\theta\|$ over $\theta \in \Theta$ and K is a random constant. The compactness of Θ implies that $\rho < 1$. We then compute

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

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

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

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

$$E\ell_t(\theta) \geq E\sum_{m=1}^{\infty} w_m \log\langle \sigma_{\theta t}^2, \varphi_m\rangle \geq \sum_{m=1}^{\infty} w_m(\log c + \log a_m) \geq \text{const} - \sum_{m=1}^{\infty} w_m \log\left(\frac{1}{a_m}\right) > -\infty,$$

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

$$\begin{aligned}
E\,\ell_t(\theta_0) &= \sum_{m=1}^{\infty} w_m E\Big\{E\Big[\frac{\langle y_t^2, \varphi_m\rangle}{\langle \sigma_t^2, \varphi_m\rangle}\Big|\mathcal{F}_{t-1}\Big] + \log\langle \sigma_t^2, \varphi_m\rangle\Big\} \\
&= \sum_{m=1}^{\infty} w_m\Big(1 + E\log\langle \sigma_t^2, \varphi_m\rangle\Big) \leq \sum_{m=1}^{\infty} w_m\Big(1 + E\log\|\sigma_{1_0}^2\|\Big) < \infty.
\end{aligned}$$

We now have that

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

with equality iff for all $m \geq 1$

$$\langle \sigma_t^2, \varphi_m \rangle = \langle \sigma_t^2, \varphi_m \rangle, \quad \text{a.s.} \tag{5.16}$$

The proof of (iii) will be completed using (ii). To show (ii) we suppose that (5.15) holds true. We then have, for all $m \geq 1$,

$$\begin{aligned} & \langle \delta_0, \varphi_m \rangle + \langle \boldsymbol{\alpha}_0(y_{t-1}^2), \varphi_m \rangle + \langle \boldsymbol{\beta}_0(\sigma_{t-1}^2), \varphi_m \rangle \\ = \; & \langle \delta_\theta, \varphi_m \rangle + \langle \boldsymbol{\alpha}_\theta(y_{t-1}^2), \varphi_m \rangle + \langle \boldsymbol{\beta}_\theta(\sigma_{t-1}^2), \varphi_m \rangle \;\; \text{a.s.} \end{aligned} \tag{5.17}$$

We have,

$$\langle \boldsymbol{\alpha}_\theta(y_{t-1}^2), \varphi_m \rangle = \langle \sigma_{t-1_0}^2 \eta_{t-1}^2, \boldsymbol{\alpha}_\theta^*(\varphi_m) \rangle = \langle \eta_{t-1}^2, \sigma_{t-1_0}^2 \boldsymbol{\alpha}_\theta^*(\varphi_m) \rangle.$$

In view of (5.17) we have

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

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

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

By **A8** (b) we deduce that $\boldsymbol{\alpha}_\theta = \boldsymbol{\alpha}_0$. From (5.17) we have, moreover, that

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

or, equivalently, that

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

It follows that

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

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

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

or, equivalently, that

$$\langle \eta_{t-2}^2, \sigma_{\theta_0 t-2}^2 (\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}_0^*)(\varphi_m) \rangle = \langle \eta_{t-2}^2, \sigma_{\theta_0 t-2}^2 (\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}_\theta^*)(\varphi_m) \rangle + K_{t-3}, \;\; \text{a.s.}$$

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

$$\|\widehat{\theta}_n^{N_n} - \theta_0\| \leq \|\widehat{\theta}_n^{N_n} - \widehat{\theta}_n|_{\Theta_{N_n}}\| + \|\widehat{\theta}_n|_{\Theta_{N_n}} - \widehat{\theta}_n\| + \|\widehat{\theta}_n - \theta_0\|, \tag{5.18}$$

where $|_{\Theta_N}$ denotes the projection $x \mapsto (x_1, \ldots, x_N)$. Up to now, we have shown that the third term of (5.18) converges almost surely to zero. The second term is equal to $\sum_{j > N_n} \widehat{\theta}_{n,j}$ 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 this quantity by $\sup_{\ell \geq 1} \sum_{j > N_n} \widehat{\theta}_{\ell,j}$. To show that this converges to zero we apply the tighness Lemma 14 in Cerovecki and Hörmann (2017) with $p_j^{(n)} = \widehat{\theta}_{n,j}$ and $p_j^{(0)} = \theta_{0,j}$. Finally, from the compactness of Θ we know that there exists a subsequence $(\widehat{\theta}_{n_\ell}^{N_{n_\ell}})_{\ell \geq 1}$ that converges in ℓ^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 θ, 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 (5.18) converges a.s. to zero. $\square$

References

Aue, A., Dubart Norinho, D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110, 378–392.

Aue, A., Horváth, L. and Pellatt., D. F. (2016). Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis*, 38, 3–21.

Aue, A. and Klepsch, J. (2017). Estimating invertible functional time series. *Functional Statistics and Related Fields*, 51–58.

Billingsley, P.(1971). *Convergence of Probability Measures.* First edition, Wiley series in probability and mathematical statistics, New York.

Bollerslev, T.(1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327.

Bougerol, P. and Picard, N. (1992). Strict stationarity of generalized autoregressive processes. *The Annals of Probability*, 20, 1714–1730.

Cerovecki, C. and Hörmann, S. (2017). On the CLT for discrete Fourier transforms of functional time series. *Journal of Multivariate Analysis*, 154, 282–295.

Eichler, M. and 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. and Zakoïan, J. M. (2011). *GARCH models: structure, statistical inference and financial applications.* John Wiley & Sons.

Hörmann, S., Horváth, L. and Reeder, R. (2013). A functional version of the ARCH model. *Econometric Theory*, 29, 267–288.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38, 1845–1884.

Horváth, L., Kokoszka, P. (2014). Inference for functional data with applications (Vol. 200). Springer Science & Business Media.

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179, 66–82.

Kingman, J. F. C. (1973). Subadditive ergodic theory. *The Annals of Probability*, 1, 883–899.

Klepsch, J. and Klüppelberg, C. (2017). An innovations algorithm for the prediction of functional linear processes. *Journal of Multivariate Analysis*, 155, 252–271.

Paparoditis, E. (2017). Sieve bootstrap for functional time series. *arXiv preprint* 1609.06029v2.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis.* Springer, New York, second edition.

Zhu, T. and Politis, D.N. (2017). Kernel estimation of first-order nonparametric functional autoregression model and its bootstrap approximation. *Electronic Journal of Statistics*, forthcoming.