Tests for conditional heteroscedasticity with functional data and goodness-of-fit tests for FGARCH models

Rice, Gregory and Wirjanto, Tony and Zhao, Yuqian

University of Waterloo, University of Waterloo, University of Waterloo

31 March 2019

TESTS FOR CONDITIONAL HETEROSCEDASTICITY WITH FUNCTIONAL DATA AND GOODNESS-OF-FIT TESTS FOR FGARCH MODELS

GREGORY RICE*, TONY WIRJANTO* AND YUQIAN ZHAO*

**Department of Statistics and Actuarial Science, University of Waterloo, Canada*

Abstract. Functional data objects that are derived from high-frequency financial data often exhibit volatility clustering characteristic of conditionally heteroscedastic time series. Versions of functional generalized autoregressive conditionally heteroscedastic (FGARCH) models have recently been proposed to describe such data, but so far basic diagnostic tests for these models are not available. We propose two portmanteau type tests to measure conditional heteroscedasticity in the squares of financial asset return curves. A complete asymptotic theory is provided for each test, and we further show how they can be applied to model residuals in order to evaluate the adequacy, and aid in order selection of FGARCH models. Simulation results show that both tests have good size and power to detect conditional heteroscedasticity and model misspecification in finite samples. In an application, the proposed tests reveal that intra-day asset return curves exhibit conditional heteroscedasticity. Additionally, we found that this conditional heteroscedasticity cannot be explained by the magnitude of inter-daily returns alone, but that it can be adequately modeled by an FGARCH(1,1) model.

JEL Classification: C12, C32, C58, G10

Keywords: Functional time series, Heteroscedasticity testing, Model diagnostic checking, High-frequency volatility models, Intra-day asset price

1. Introduction

Since the seminal work of Engle (1982) and Bollerslev (1986), generalized autoregressive conditionally heteroscedastic (GARCH) models and their numerous generalizations have become a cornerstone of financial time series modeling, and are frequently used as a model for the volatility of financial asset returns. As the name suggests, the main feature that these models account for is conditional heteroscedasticity, which for an uncorrelated financial time series can be detected by checking for the presence of serial correlation in the series of squared returns of the asset. This basic observation leads to several ways of testing for the presence of conditional heteroscedasticity in a given time series or series of model residuals by applying portmanteau tests to the squared series. Such tests have been developed by McLeod and Li (1983) and Li and Mak (1994) to test for conditional heteroscedasticity and perform model selection for GARCH models as well as autoregressive moving average models with GARCH errors. Diagnostic tests of this type are summarized in the monograph by Li (2003), and with a special focus on GARCH models in Francq and Zakoïan (2010). Many of these methods have also been extended to multivariate time series of a relatively small dimension; see also Francq and Zakoïan (2010), Tse and Tsui (1999), Tse (2002), Duchesne and Lalancette (2003), Kroner and Ng (1998), Bauwens *et al.* (2006), and Catani *et al.* (2017).

In many applications, dense intra-day price data of financial assets are available in addition to the daily asset returns. One way to view such data is as daily observations of high dimensional vectors (consisting of hundreds or thousands of coordinates) that may be thought of as discrete observations of an underlying noisy intra-day price curve or function. We illustrate with the data that motivate our work and will be further studied below. On consecutive days $i \in \{1, \dots, N\}$, observations of the price of an asset, for instance the index of Standard & Poor's 500, are available at intra-day times u, measured at a 1-minute (or finer) resolution. These data may then be represented by a sequence of discretely observed functions $\{P_i(u) : 1 \le i \le T, u \in [0, S]\}$, with S denoting the length of the trading day. Transformations of these functions towards stationarity that are of interest include the horizon h log returns, $R_i(u) = \log P_i(u) - \log P_i(u - h)$, where h is some given length of time, such as five minutes. For a fixed h, on any given trading day i, we thus observe a high-dimensional multivariate vector that can be viewed as a curve. The collection of these curves can therefore be studied as a functional time series. We refer the reader to Bosq (2000), Ramsay and Silverman (2006), and Horváth and Kokoszka (2012) for a review of functional data analysis and linear functional time series. Studying such data through the lens of a functional data analysis has received considerable attention in recent years. The basic idea of viewing transformations of densely observed asset price data as sequentially observed stochastic processes appears in studies such as Barndorff-Nielson and Shepard (2004), Müller *et al.* (2011) and Kokoszka and Reimherr (2013), among others.

Curves produced as described above exhibit a non-linear dependence structure and volatility clustering reminiscent of GARCH-type time series. Recently functional GARCH (FGARCH) models have been put forward as a model for curves derived from the dense intra-day price data, beginning with Hörmann *et al.* (2013), who proposed an FARCH(1) model, which was generalized to FGARCH(1,1) and FGARCH(p, q) models by Aue *et al.*, (2017), and Cerovecki *et al.* (2019), respectively. An important determination an investigator may wish to make before she employs such a model is whether or not the observed functional time series exhibits substantial evidence of conditional heteroscedasticity. To the best of our knowledge, there is no formal statistical test available to measure conditional heteroscedasticity in intra-day return curves or generally for sequentially observed functional data. Additionally, if an FGARCH model is employed, it is desirable to know how well it fits the data, and whether or not the orders p and q selected for the model should be adjusted. This can be addressed by testing for remaining conditional heteroscedasticity in the model residuals of fitted models.

In this paper, we develop functional portmanteau tests for the purpose of identifying conditional heteroscedasticity in functional time series. Additionally, we consider applications of the proposed tests to the model residuals from a fitted FGARCH model that can be used to evaluate the model's adequacy and aid in the order selection. The development of this later application entails deriving joint asymptotic results between the autocovariance of the FGARCH innovations and the model parameter estimators that are of independent interest. Simulation studies presented in this paper confirm that the proposed tests have good size and are effective in identifying functional conditional heteroscedasticity as well as mis-specification of FGARCH-type models. In an application to intra-day return curves derived from dense stock price data, our tests suggest that the FGARCH models are adequate for modeling the observed conditional heteroscedasticity across curves.

This work builds upon a number of recent contributions related to portmanteau and goodness-of-fit tests for functional data. Gabrys and Kokoszka (2007) were the first to consider white noise tests for functional time series, and their initial approach was based on portmanteau statistics applied to finite-dimensional projections of functional observations. Horváth *et al.* (2013) developed a general strong white noise test based on the squared norms of the autocovariance operators for an increasing number of lags. General weak white noise tests that are robust to potential conditional heteroscedasticity were developed in Zhang (2016) and Kokoszka *et al.* (2017). Zhang (2016), Gabrys *et al.* (2010) and Chiou and Müller (2007) also consider goodness-of-fit tests based on model residuals, with the first two being in the context of modeling functional time series.

The remainder of the paper is organized as follows. In Section 2 we frame testing for conditional heteroscedasticity as a hypothesis testing problem, and introduce test statistics for this purpose. We further present the asymptotic properties of the proposed statistics, and show how to apply them to the model residuals of the FGARCH models for the purpose of model validation/selection. Some details regarding the practical implementation of the proposed tests and a simulation study evaluating their performance in finite samples are given in Section 4. An application to intra-day return curves is detailed in Section 5, and concluding remarks are made in Section 6. Proofs of the asymptotic results are collected in appendices following these main sections.

We use the following notation below. We let $L^2[0,1]^d$ denote the space of real valued square integrable functions defined on unit hypercube $[0,1]^d$ with norm $\|\cdot\|$ induced by the inner product $\langle x, y\rangle = \int_0^1 \cdots \int_0^1 x(t_1,...,t_d)y(t_1,...,t_d)dt_1 \ldots dt_d$ for $x, y \in L^2[0,1]^d$, the dimension of the domain being clear based on the input function. Henceforth we write $\int$ instead of $\int_0^1$. We often consider kernel integral operators of the form $\boldsymbol{g}(x)(t) = \int g(t,s)x(s)ds$ for $x \in L^2[0,1]$, where the kernel function g is an element of $L^2[0,1]^2$. We use $\boldsymbol{g}^{(k)}(x)(t)$ to denote the k-fold convolution of the operator $\boldsymbol{g}$. The filtration $\mathcal{F}_i$ is used to denote the sigma algebra generated by the random elements $\{X_j,\ j \le i\}$. We let $C[0,1]$ denote the space of continuous real valued functions on $[0,1]$, with norm defined for $x \in C[0,1]$ as $\|x\|_\infty = \sup_{y\in[0,1]} |x(y)|$. We let χ^2_K denote a chi-square random variable with K degrees of freedom, and use $\chi^2_{K,q}$ to denote its q'th quantile. $\|\cdot\|_E$ denotes the standard Euclidean norm of a vector in $\mathbb{R}^d$. We use $\{x_i\}$ to denote the sequence $\{x_i\}_{i\in\mathbb{N}}$, or $\{x_i\}_{i\in\mathbb{Z}}$, with the specific usage of which being clear in context.

2. Tests for functional conditional heteroscedasticity

Consider a stretch of a functional time series of length N, $X_1(t), ..., X_N(t)$, which is assumed to have been observed from a strictly stationary sequence $\{X_i(t),\ i\in\mathbb{Z},\ t\in[0,1]\}$ of stochastic processes with sample paths in $L^2[0,1]$. For instance, below $X_i(t)$ denotes the intra-day log returns derived from densely observed stock prices on day i at intraday time t where t is normalized to be in the unit interval. In this paper, we are generally concerned with developing tests that differentiate such series of curves, or model residuals, exhibiting conditional heteroscedasticity from those that are strong functional white noises.

As emphasized by Engle (1982), conditional heteroscedasticity is generally characterized by dependence of the conditional variance of an observed scalar time series on the magnitude of its past values, which manifests itself in serial correlation in the squares of the series. This leads one to consider the following definition of conditional heteroscedasticity for functional observations:

Definition 2.1. [Functional Conditional Heteroscedasticity] We say that a sequence $\{X_i\}$ is conditionally heteroscedastic in $L^2[0,1]$ if it is strictly stationary, $E[X_i(t)|\mathcal{F}_{i-1}]=0$, and

$$\mathrm{cov}(X_i^2(t), X_{i+h}^2(s)) \neq 0,$$

for some $h\geq 1$, where the equality above is understood to be in the $L^2[0,1]^2$ sense.

Recently, several models have been proposed in order to model series of curves exhibiting conditional heteroscedasticity. Notably, the functional ARCH(1) and GARCH(1,1) processes were put forward by Hörmann *et al.* (2013) and Aue *et al.* (2017), respectively, and take the form

$$X_i(t)=\sigma_i(t)\varepsilon_i(t),\quad E\varepsilon_i^2(t)=1,\ \ t\in[0,1], \tag{2.1}$$

where

$$\text{FARCH}(1):\ \ \sigma_i^2(t)=\omega(t)+\boldsymbol{\alpha}(X_{i-1}^2)(t)=\omega(t)+\int\alpha(t,s)X_{i-1}^2(s)ds, \tag{2.2}$$

or FGARCH$(1,1)$:

$$\sigma_i^2(t) = \omega(t) + \boldsymbol{\alpha}(X_{i-1}^2)(t) + \boldsymbol{\beta}(\sigma_{i-1}^2)(t) = \omega(t) + \int \alpha(t,s)X_{i-1}^2(s)ds + \int \beta(t,s)\sigma_{i-1}^2(s)ds,$$

respectively. Here $\omega(t)$ is a non-negative intercept function, and $\alpha(t,s)$ and $\beta(t,s)$ are non-negative kernel functions. General FGARCH(p,q) models are discussed in Cerovecki *et al.* (2019), in which they also provide natural conditions under which these models admit strictly stationary and non-anticipative solutions.

We frame testing for conditional heteroscedasticity as a hypothesis testing problem of

$\mathcal{H}_0$: The sequence $\{X_i(t)\}$ is independent, and identically distributed, versus

$\mathcal{H}_A$: The sequence of $\{X_i(t)\}$ is conditionally heteroscedastic given in Definition 2.1.

Clearly it is not the case in general that rejecting $\mathcal{H}_0$ would directly lead us to $\mathcal{H}_A$, because $X_i(t)$ might instead be dependent or corrleated in the first moment. This concern can be alleviated though if we test serial correlation in the sequence of squared curves as described in Definition 2.1.

In particular, we might then test $\mathcal{H}_0$ versus $\mathcal{H}_A$ by measuring the serial correlation in the time series $\|X_1\|^2$,...,$\|X_N\|^2$, or in the sequence of curves $X_1^2(t)$,...,$X_N^2(t)$. Testing for serial correlation in the time series $\|X_i\|^2$ can be viewed as measuring to what extent large in magnitude curves increase/decrease the likelihood of subsequent curves being large in magnitude, whereas testing for serial correlation in the curves $X_i^2(t)$ aims to more directly evaluate whether the data follow Definition 2.1. For some positive integer K, we then consider portmanteau statistics of the form

$$V_{N,K} = N\sum_{h=1}^{K}\hat{\rho}_h^2, \text{ and } M_{N,K} = N\sum_{h=1}^{K}\|\hat{\gamma}_h\|^2\,, \tag{2.3}$$

where $\hat{\rho}_h$ is the sample autocorrelation of the time series $\|X_1\|^2$,...,$\|X_N\|^2$, and $\hat{\gamma}_h(t,s) \in L^2[0,1]^2$ is the estimated autocovariance kernel of the functional time series $X_i^2(t)$ at lag h , defined as

$$\hat{\gamma}_h(t,s) = \frac{1}{N}\sum_{i=1}^{N-h}(X_i^2(t) - \bar{X}^{(2)}(t))(X_{i+h}^2(s) - \bar{X}^{(2)}(s)),$$

with $\bar{X}^{(2)}(t)) = (1/N)\sum_{i=1}^{N} X_i^2(t)$. The test statistic $V_{N,K}$ is essentially the Box-Ljung-Pierce test statistic (Ljung and Box, 1978) derived from the scalar time series of squared norms, whereas the test statistic $M_{N,K}$ is the same as the portmanteau statistic defined in Kokoszka *et al*. (2017) applied to the squared functions.

Under $\mathcal{H}_A$, we expect the statistics $V_{N,K}$ and $M_{N,K}$ to be large, and hence a consistent test can be obtained by rejecting $\mathcal{H}_0$ whenever they exceed a threshold calibrated according to their limiting distributions under the null hypothesis. In order to establish the asymptotic distributions of each portmanteau statistic under $\mathcal{H}_0$, we impose the following moment condition.

Assumption 2.1. *$E\|X_i\|^8 < \infty$, $i \in \mathbb{Z}$.*

Under this assumption, the asymptotic distribution of $M_{N,K}$ depends on the eigenvalues $\lambda_i,\ i \geq 1$ of the kernel integral operator with kernel $\text{cov}(X_i^2(t), X_i^2(s))$, namely

$$\lambda_i \varphi_i(t) = \int \text{cov}(X_i^2(t), X_i^2(s))\varphi_i(s)ds, \tag{2.4}$$

where $\{\varphi_i\}$ is an orthonormal sequence of eigenfunctions in $L^2[0,1]$. Assumption 2.1 guarantees that the eigenvalues $\{\lambda_i\}$ satisfy the condition that $\sum_{i=1}^{\infty} \lambda_i < \infty$.

Theorem 2.1. *If $\mathcal{H}_0$ and Assumption 2.1 are satisfied, then we have*

$$V_{N,K} \xrightarrow{\mathcal{D}} \chi^2_K, \text{ as } N \to \infty, \tag{2.5}$$

and

$$M_{N,K} \xrightarrow{\mathcal{D}} \sum_{h=1}^{K} \sum_{l,k=1}^{\infty} \lambda_l \lambda_k \chi_1^2(h, \ell, k), \text{ as } N \to \infty, \tag{2.6}$$

where $\{\chi_1^2(h,\ell,k),\ 1 \leq h \leq K,\ 1 \leq \ell, k < \infty\}$ are independent and identically distributed χ_1^2 random variables.

Theorem 2.1 shows that an approximate test of H_0 of size q is to reject if $V_{N,K} > \chi^2_{K,1-q}$ or if $M_{N,K}$ exceeds the q'th quantile of the distribution on the right hand side of (2.6). The latter can be approximated in several ways, and in Section 4 below we describe a Welch-Satterthwaite style χ^2 approximation to achieve this.

2.1. **Consistency of the proposed tests.** We now turn to studying consistency of each test under $\mathcal{H}_A$. In particular, we consider the asymptotic behavior of $V_{N,K}$ and $M_{N,K}$ for sequences $\{X_i\}$ such that either: (a) they form a general weakly dependent sequences in $L^2[0,1]$ that are conditionally heteroscedastic as described by Definition 2.1, or (b) they follow a FARCH(1) model as described in (2.2). We use the notion of L^p-m-approximability defined in Hörmann and Kokoszka (2010) in order to describe general weakly dependent sequences, which covers strictly stationary functional GARCH type processes under suitable moment conditions; see Cerovecki *et al.* (2019).

Theorem 2.2. *If $\{X_i\}$ is L^8-m-approximable and $\mathcal{H}_A$ holds where h in Definition 2.1 satisfies $1 \le h \le K$, then*

$$M_{N,K} \xrightarrow{p} \infty, \quad N \to \infty. \tag{2.7}$$

If in addition $\iint cov(X_i^2(t), X_{i+h}^2(s))dtds \neq 0$, then

$$V_{N,K} \xrightarrow{p} \infty, \quad N \to \infty. \tag{2.8}$$

Remark 2.1. In typical financial applications we expect that the sequence of squared returns are positively correlated, which may be interpreted in this setting as $\text{cov}(X_i^2(t), X_{i+h}^2(s)) \ge 0$, for all $t, s \in [0,1]$, *i.e.* the covariance surface of the squared process at lag h of $X_i^2(t)$ is strictly positive. Under this additional requirement the conditions for consistency of $M_{N,K}$ and $V_{N,K}$ in Theorem 2.2 become equivalent.

Under the FARCH(1) model we can develop more precise results on the rate of divergence of $V_{N,K}$ and $M_{N,K}$. The following assumption ensures that a stationary and causal sequence satisfying (2.1) and (2.2) exists in $L^2[0,1]$:

Assumption 2.2. *The sequence $\{\varepsilon_i\}$ in* (2.1) *is independent and identically distributed, and the kernel $\alpha(t,s)$ in* (2.2) *is non-negative, $\|\alpha\| < 1$, and satisfies that there exists a constant $\tau > 0$ so that*

$$E\left(\iint \alpha^2(t,s)\varepsilon_0^2(s)dtds\right)^{\tau/2} < 1.$$

Theorem 2.3. *Suppose that $\{X_i\}$ is the strictly stationary solution to the FARCH(1) equations under Assumption 2.2 so that Assumption 2.1 holds, and let $V_i(t) = X_i^2(t) - \sigma_i^2(t)$. Then $V_i(t)$ is a mean zero weak white noise in $L^2[0,1]$ (see pg. 72 Bosq (2000)),*

$$\frac{V_{N,K}}{N} \xrightarrow{p} \sum_{h=1}^{K} \frac{\left(\iint \sum_{j=0}^{\infty} E\boldsymbol{\alpha}^{(j)}(V_j)(t)\boldsymbol{\alpha}^{(j+h)}(V_j)(s)dtds\right)^2}{\left(\iint \sum_{j=0}^{\infty} E\boldsymbol{\alpha}^{(j)}(V_j)(t)\boldsymbol{\alpha}^{(j)}(V_j)(s)dtds\right)^2}, \tag{2.9}$$

and

$$\frac{M_{N,K}}{N} \xrightarrow{p} \sum_{h=1}^{K} \left\| \sum_{j=0}^{\infty} E\boldsymbol{\alpha}^{(j)}(V_j)(t)\boldsymbol{\alpha}^{(j+h)}(V_j)(s) \right\|^2. \tag{2.10}$$

The right hand side of (2.10) *is guaranteed to be strictly positive if* $\iint \alpha(t,s)E\omega(t)(\varepsilon_0^2(t) - 1)\omega(s)(\varepsilon_0^2(s) - 1)dtds \neq 0$.

Remark 2.2. Theorem 2.3 shows that under an FARCH(1) model, the rate of divergence of $V_{N,K}$ and $M_{N,K}$ depend essentially on the size of the function $\alpha(t,s)$ as well as how this kernel projects onto the intercept term in the conditional variance $\omega(t)$ and the covariance of the squared error $\varepsilon_0^2(t)$. If for example $\iint \alpha(t,s)E(\varepsilon_0^2(t) - 1)(\varepsilon_0^2(t) - 1)dtds = 0$, then we do not expect the tests to be consistent.

3. Diagnostic Checking for Functional GARCH Models

The conditional heteroscedasticity tests proposed above can also be used to test for the adequacy of the estimated functional ARCH and GARCH models, and can aid in the order selection of these models. We introduce this approach in the context of testing the adequacy of the FGARCH(1,1) model, although one could more generally consider the same procedure applied to the FGARCH(p,q) models using the estimation procedures in Cerovecki *et al*. (2019). To this end, suppose that $X_i(t)$, $1 \leq i \leq N$ follows an FGARCH(1,1) model. To estimate $\omega(t)$, and the kernel functions $\alpha(t,s)$ and $\beta(t,s)$, following Aue *et al*. (2017) and Cerovecki *et al*. (2019), we suppose that they have finite L-dimensional representations determined by a set of basis functions $\Phi_L = \{\phi_1, \phi_2, \dots, \phi_L\}$ in $L^2[0,1]$ so that

$$\text{(3.1)} \qquad \omega(t) = \sum_{j=1}^{L} d_j \phi_j(t), \alpha(t,s) = \sum_{j,j'=1}^{L} a_{j,j'} \phi_j(t) \phi_{j'}(s), \beta(t,s) = \sum_{j,j'=1}^{L} b_{j,j'} \phi_j(t) \phi_{j'}(s).$$

Under this assumption, estimating these functions amounts to estimating the coefficients in their finite dimensional representations, which can be achieved by using, for example, Quasi-Maximum Likelihood estimation (QMLE) or Least Squares estimation, as is typically employed in multivariate GARCH models. To see this, under (3.1) we can re-express the FGARCH(1,1) model in terms of the coefficients as

$$\text{(3.2)} \qquad \mathfrak{s}_i^2 = D + A\mathbf{x}_{i-1}^2 + B\mathfrak{s}_{i-1}^2$$

where $\mathbf{x}_i^2 = [\langle X_i^2(t), \phi_1(t) \rangle, \ldots, \langle X_i^2(t), \phi_L(t) \rangle]^\top$, $\mathfrak{s}_i^2 = [\langle \sigma_i^2(t), \phi_1(t) \rangle, \ldots, \langle \sigma_i^2(t), \phi_L(t) \rangle]^\top$, the coefficient vector $D = [d_1, \ldots, d_L]^\top \in \mathbb{R}^L$, and the coefficient matrices A and B are $\mathbb{R}^{L \times L}$ with (j, j') entries by $a_{j,j'}$ and $b_{j,j'}$, respectively. To estimate the vector of parameters $\theta_0 = (D^\top, \text{vec}(A)^\top, \text{vec}(B)^\top)^\top$, Aue *et al.* (2017) propose a Least Squares type estimator satisfying

$$\hat{\theta}_N = \arg\min_{\theta \in \boldsymbol{\Theta}} \left\{ \sum_{i=2}^{N} (\mathbf{x}_i^2 - \mathfrak{s}_i^2(\theta))^\top (\mathbf{x}_i^2 - \mathfrak{s}_i^2(\theta)) \right\},$$

where $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^{L+2L^2}$. Under certain regularity conditions, detailed at the beginning of Appendix B, it can be shown that $\hat{\theta}_N$ is a consistent estimator of θ_0, and in fact $\sqrt{N}(\hat{\theta}_N - \theta_0)$ satisfies the central limit theorem. This yields estimated functions given by

$$\hat{\omega}(t) = \sum_{j=1}^{L} \hat{d}_j \phi_j(t), \quad \hat{\alpha}(t,s) = \sum_{j,j'=1}^{L} \hat{a}_{j,j'} \phi_j(t) \phi_{j'}(s), \quad \hat{\beta}(t,s) = \sum_{jj'=1}^{L} \hat{b}_{j,j'} \phi_j(t) \phi_{j'}(s).$$

The functions ϕ_j can be chosen in a number of ways, including using a deterministic basis system such as polynomials, b-splines, or the Fourier basis, as well as using a functional principal component basis; see *e.g.* Chapter 6 of Ramsay and Silverman (2006). Cerovecki *et al.* (2019) and Aue *et al.* (2017) suggest using the principal component basis determined by the squared processes $X_i^2(t)$, which we also consider below. Given these function estimates, we can estimate recursively $\hat{\sigma}_i^2(t)$, see (B.4) in Appendix B for specific details.

To test the adequacy of the FGARCH(1,1) model, we utilize the fact that if the model is well specified then the sequence of model residuals $\varepsilon_i(t)$, $1 \leq i \leq N$, should be approximately independent and identically distributed, where

$$\hat{\varepsilon}_i(t) = \frac{X_i(t)}{\hat{\sigma}_i(t)}. \tag{3.3}$$

This suggests that we consider the portmanteau statistics constructed from the residuals

$$V_{N,K,\varepsilon} = N \sum_{h=1}^{K} \hat{\rho}_{\varepsilon,h}, \text{ and } M_{N,K,\varepsilon} = N \sum_{h=1}^{K} \|\hat{\gamma}_{\varepsilon,h}\|^2,$$

where $\hat{\rho}_{\varepsilon,h}$ is the sample autocorrelation of the scalar time series $\|\hat{\varepsilon}_1\|^2, \ldots, \|\hat{\varepsilon}_N\|^2$, and

$$\hat{\gamma}_{\varepsilon,h}(t,s) = \frac{1}{N} \sum_{i=1}^{N-h} \left(\hat{\varepsilon}_i^2(t) - 1\right) \left(\hat{\varepsilon}_{i+h}^2(s) - 1\right). \tag{3.4}$$

A test of model adequacy of size q is to reject if $V_{N,K,\varepsilon} > \chi^2_{K,1-q}$ or if $M_{N,K,\varepsilon}$ exceeds the $1-q$'th quantile of the distribution on the right hand side of (2.6), where again this distribution must be estimated from the squared residuals $\hat{\varepsilon}_i^2(t)$. We abbreviate these tests below as being based on $V_{N,K,\varepsilon}^{heuristic}$ and $M_{N,K,\varepsilon}^{heuristic}$, since even under the assumption that that the model is correctly specified the residuals $\hat{\varepsilon}_i$ are evidently not independent and identically distributed due to their common dependence on the estimated parameters $\hat{\theta}_N$.

3.1. **Accounting for the effect of parameter estimation.** The approximate goodness-of-fit tests proposed above provide a heuristic method to evaluate the model fit of a specified functional GARCH type model, however we now aim at more precisely describing how the asymptotic distribution of $M_{N,K,\varepsilon}$ based on the model residuals $\hat{\varepsilon}_i(t)$ depends on the joint asymptotics of the innovation process and the estimated parameters $\hat{\theta}_N$. In this subsection, we focus only on quantifying this effect for the fully functional statistic $M_{N,K,\varepsilon}$. Further, we assume that the parameter estimate $\hat{\theta}_N$ is obtained by the Least Squares method proposed in Aue et al. (2017), although this could easily be adapted to the QMLE parameter estimate as well.

Given the regularity conditions stated Appendix B, it follows that

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} \mathcal{N}_{L+2L^2}(0, Q_0^{-1} H_0^\top J_0 H_0 Q_0^{-1}), \tag{3.5}$$

where $\mathcal{N}_p(0, \Sigma)$ denotes a p dimensional normal random vector with mean zero and covariance matrix Σ. We use the notation $\sigma_i^2(t, \theta)$ and $\mathfrak{s}_i^2(\theta)$ to indicate how each of these terms depends on the vector of parameters defined in (3.1). The terms J_0, H_0, and Q_0 are respectively defined as

$$J_0 = E\{[\mathbf{x}_0^2 - \mathfrak{s}_0^2][\mathbf{x}_0^2 - \mathfrak{s}_0^2]^\top\}, \quad H_0 = E\left\{\frac{\partial \tilde{\mathfrak{s}}_0^2(\theta)}{\partial \theta}\right\}, \quad Q_0 = E\left\{\left[\frac{\partial \tilde{\mathfrak{s}}_0^2(\theta)}{\partial \theta}\right]^\top \left[\frac{\partial \tilde{\mathfrak{s}}_0^2(\theta)}{\partial \theta}\right]\right\}.$$

Let $G_h : [0,1]^2 \to \mathbb{R}^{L+2L^2}$ be defined by

$$G_h(t, s) = -E\left\{\frac{1}{\sigma_{i+h}^2(s, \theta_0)} \times \frac{\partial \sigma_{i+h}^2(s, \theta_0)}{\partial \theta} \times (\varepsilon_i^2(t, \theta_0) - 1)\right\}. \tag{3.6}$$

We further define the covariance kernels

$$C_\varepsilon(t, s, u, v) = E\{(\varepsilon_i^2(t) - 1)(\varepsilon_i^2(s) - 1)\}E\{(\varepsilon_i^2(u) - 1)(\varepsilon_i^2(v) - 1)\},$$

and

$$C_{h,g}^{\varepsilon,\theta}(t, s, u, v) = E\left\{(\varepsilon_{-h}^2(t) - 1)(\varepsilon_0^2(s) - 1)G_g^\top(u, v)Q_0^{-1}(\frac{\partial \mathfrak{s}_0^2(\theta_0)}{\partial \theta})^\top(\mathbf{x}_0^2 - \mathfrak{s}_0^2)\right\}.$$

Theorem 3.1. *Suppose that $\{X_i\}$ follows an FGARCH(1,1) model. Under the assumptions detailed in Appendix B, there exists a sequence of non-negative coefficients $\{\xi_{i,K}^{(\varepsilon,\theta)}, \quad i \geq 1\}$ such that*

$$M_{N,K,\varepsilon} \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty} \xi_{i,K}^{(\varepsilon,\theta)} \chi_1^2(i), \tag{3.7}$$

where $\chi_1^2(i),\ i \geq 1$ are independent and identically distributed χ^2 random variables with one degree of freedom. The coefficients $\xi_{i,K}^{(\varepsilon,\theta)}$ are the eigenvalues of a covariance operator $\Psi_K^{(\varepsilon,\theta)}$, defined in (B.1) *below, that is constructed from kernels of the form*

$$\begin{aligned}\psi_{K,h,g}^{(\varepsilon,\theta)}(t, s, u, v) &= C_\varepsilon(t, s, u, v) + C_{h,g}^{\varepsilon,\theta}(t, s, u, v) \\ &\quad + C_{g,h}^{\varepsilon,\theta}(u, v, t, s) + G_h^\top(t, s)Q_0^{-1}H_0^\top J_0 H_0 Q_0^{-1} G_g(u, v),\ 1 \leq h, g \leq K.\end{aligned} \tag{3.8}$$

Theorem 3.1 more precisely details the asymptotics for $M_{N,K,\varepsilon}$, which in this case depend jointly on the autocovariance of the FGARCH innovations as well as the parameter estimates. A rigorous statement of this result is given in Appendix B along with the necessary assumptions on the FGARCH model, which basically are taken to be strong enough to imply (3.5), and that the solution $\{X_i\}$ of the FGARCH equations exists in $C[0,1]$ with sufficient moments. These results may be easily generalized to FGARCH models of other orders, for instance, the FARCH(1) model, which we study in the simulation section below.

4. Implementation of the tests and a simulation study

This section gives details on implementation of the proposed tests and evaluates the performance of the proposed tests in finite samples. Several synthetic data examples are considered for this purpose. A simulation study on diagnostic checking for the FGARCH model is also provided in the last subsection.

4.1. **Computation of test statistics and asymptotic critical values.** In practice we only observe each functional data object $X_i(t)$ at a discrete collection of time points. Often in financial applications these time points can be taken to be regularly spaced and represented as $\mathcal{T}_J = \{t_j = j/J,\ j = 1, \ldots, J\} \subset (0,1]$. Given the observations of the function $X_i(t_j)$, $t_j \in \mathcal{T}_j$, we can estimate, *e.g.* the squared norm $\|X_i\|^2$ by a simple Riemann sum,

$$\|X_i\|^2 = \frac{1}{J}\sum_{j=1}^{J} X_i^2(t_j).$$

Other norms arising in the definitions of $V_{N,K}$ and $M_{N,K}$ can be approximated similarly. For data observed at different frequencies, such as tick-by-tick, the norms and inner-products can be estimated with Riemann sums or alternate integration methods as the data allows. In all of the simulations below we generate functional observations on $J = 50$ equally spaced points in the interval $[0,1]$.

The critical values of the null limiting distribution of $V_{N,K}$ can easily be obtained, but estimating the limiting null distribution of $M_{N,K}$ defined in (2.6) requires a further elaboration. One option is to directly estimate the eigenvalues of the kernel integral operator with kernel $\text{cov}(X_i^2(t), X_i^2(s))$ via estimates of the kernel. Here, for the sake of computational efficiency, we propose a Welch-Satterthwaite style approximation of the limiting distribution; see *e.g.* Zhang (2013) and Kokoszka *et al.* (2017). The basic idea of this method is to approximate the limiting distribution in (2.6) by a random variable $R_K \sim \beta\chi^2_\nu$, where β and ν are estimated so that the distribution of R_K has the same first two moments as the limiting distribution on the right hand side of (2.6). If M_K denotes the random variable on the right hand side of (2.6), $\mu_K = \mathrm{E}(M_K)$, and $\sigma_K^2 = \mathrm{var}(M_K)$, then in order that the first two moments of R_K match those of M_K we take

$$\beta = \frac{\sigma_K^2}{2\mu_K} \quad \text{and} \quad \nu = \frac{2\mu_K^2}{\sigma_K^2}. \tag{4.1}$$

We verify below that

$$\begin{aligned} \mu_K &= K\left(\int \mathrm{cov}(X_0^2(t), X_0^2(t))dt\right)^2, \\ \sigma_K^2 &= 2K\left(\iint \mathrm{cov}(X_0^2(t), X_0^2(s))dtds\right)^2. \end{aligned} \tag{4.2}$$

These can be consistently estimated by

$$\hat{\mu}_K = K\left(\int \frac{1}{N}\sum_{i=1}^{N}(X_i^2(t) - \bar{X}^{(2)}(t))^2 dt\right)^2, \text{ and}$$

$$\hat{\sigma}_K^2 = 2K\left(\int \frac{1}{N}\sum_{i=1}^{N}(X_i^2(t) - \bar{X}^{(2)}(t))(X_i^2(s) - \bar{X}^{(2)}(s))dtds\right)^2,$$

where $\bar{X}^{(2)}(t) = (1/N)\sum_{i=1}^{N} X_i^2(t)$. A test of $\mathcal{H}_0$ with an approximate size of q is to reject if $M_{N,K}$ exceeds the $1-q$ quantile of the distribution of $R_K \sim \hat{\beta}\chi^2_{\hat{\nu}}$.

Similarly, in order to estimate the asymptotic critical values of $M_{N,K,\varepsilon}$ under the FGARCH model adequacy described in Theorem 3.1, we obtain the parameters β and ν of approximated distribution by estimating,

$$\begin{aligned} \mu_K &= \text{Trace}(\Psi_K^{(\varepsilon,\theta)}), \\ \sigma_K^2 &= 2\text{Trace}([\Psi_K^{(\varepsilon,\theta)}]^2). \end{aligned} \tag{4.3}$$

We can consistently estimate these terms using estimators of the form,

$$\hat{\mu}_K = \sum_{h=1}^{K} \iint \hat{\psi}_{K,h,h}^{(\varepsilon,\theta)}(t,s,t,s)dtds, \text{ and}$$

$$\hat{\sigma}_K^2 = \sum_{h,g=1}^{K} 2 \iiiint [\hat{\psi}_{K,h,g}^{(\varepsilon,\theta)}(t,s,u,v)]^2 dvdudsdt,$$

where $\hat{\psi}_{K,h,g}^{(\varepsilon,\theta)}$ are consistent estimators of the kernels $\psi_{K,h,g}^{(\varepsilon,\theta)}$, which we define in the last subsection of Appendix B.

Calculating and storing such kernels, which can be thought of as 4-dimensional tensors, is computationally intractable if J is large, which is commonly the case when considering high-frequency financial data. For example, J=390 when using 1-minute resolution US stock market data. To solve this problem, we use a Monte Carlo integration to calculate the integrals above based on a randomly sparsified sample, with the sparse points J^* determined by drawing from a uniform distribution on $[0,1]$. Below we use $J^* = 20$ points to estimate these integrals, which seems to work well in practice.

4.2. **Simulation study of tests for conditional heteroscedasticity.** In this subsection we present the results of a simulation study in which we evaluate the proposed tests for functional conditional heteroscedasticity applied to simulated data sets. In particular, we consider the following data generating processes (DGPs). Let $\{W_i(t),\ t \in [0,\infty),\ i \in \mathbb{Z}\}$ denote independent and identically distributed sequences of standard Brownian motions. We let $\{\varphi_i(t),\ \ t \in [0,1],\ \ i \in \mathbb{N}\}$ denote the standard Fourier basis. We then consider the following five DGPs:

(a) IID-BM: $X_i(t) = W_i(t)$

(b) FARCH(1): $X_i(t)$ satisfies the FARCH(1) specification, with

$$\boldsymbol{\alpha}(x)(t) = \int 12t(1-t)s(1-s)x(s)ds,$$

and $\omega = 0.01$ (a constant function), and the innovation sequence $\varepsilon_i(t)$ follows an Ornstein-Uhlenbeck process, which is also used in other FGARCH-type processes throughout the paper:

$$\varepsilon_i(t) = e^{-t/2}W_i(e^t), \quad , \quad t \in [0,1]. \tag{4.4}$$

(c) FGARCH(1,1): $X_i(t)$ satisfies the FGARCH(1,1) specification, with

$$\boldsymbol{\alpha}(x)(t) = \int 12t(1-t)s(1-s)x(s)ds, \quad \boldsymbol{\beta}(x)(t) = \int 12t(1-t)s(1-s)x(s)ds,$$

$\omega = 0.01$ (a constant function), and $\varepsilon_i(t)$ follows (4.4).

(d) Pointwise (PW) GARCH(1,1): $X_i(t)$ follows (2.1) with

$$\sigma_i^2(t) = \omega(t) + \alpha(t)X_{i-1}^2(t) + \beta(t)\sigma_{i-1}^2(t)$$

where $\alpha(t) = (t - 0.5)^2 + 0.1$ and $\beta(t) = (t - 0.5)^2 + 0.4$.

(e) FGARCH-BEKK model: $X_i(t)$ satisfies

$$X_i(t) = \boldsymbol{\sigma}_i(\varepsilon_i)(t), \tag{4.5}$$

where $\boldsymbol{\sigma}_i(\cdot)(t)$ is a linear operator with a kernel function $\sigma_i(t, s)$, with

$$\sigma_i(t, s) = \sum_{\ell,j=1}^{2} H_i(\ell, j)\varphi_\ell(t)\varphi_j(s),$$

and

$$\varepsilon_i(t) = \sum_{\ell=1}^{2} Z_{i,\ell}\varphi_\ell(t), \quad Z_{i,\ell} \overset{iid}{\sim} \mathcal{N}(0, 1).$$

The matrix H_i follows a BEKK multivariate GARCH specification

$$H_i^2 = C^\top C + A\boldsymbol{\xi}_{i-1}\boldsymbol{\xi}_{i-1}^\top A^\top + BH_{i-1}^2 B^\top, \tag{4.6}$$

with

$$C = \begin{bmatrix} 1 & 0.3 \\ 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 0.3 & 0.01 \\ 0.01 & 0.3 \end{bmatrix}, \text{ and } B = \begin{bmatrix} 0.9 & 0.01 \\ 0.01 & 0.9 \end{bmatrix}.$$

The process IID-BM satisfies $\mathcal{H}_0$, while the remaining processes satisfy $\mathcal{H}_A$. The specific form of the FARCH and FGARCH processes are inspired by Aue *et al.* (2017) and produce sample paths that mimic high-frequency intraday returns. The FGARCH-BEKK process is meant to model the situation in which the vector valued time series obtained by projecting the functional series into a finite dimensional space satisfies a multivariate GARCH specification; see Engle and Kroner (1995) and Francq and Zakoïan (2010). The existence of a stationary and causal solution in $L^2[0, 1]$ to (4.5) follows if the multivariate GARCH specification in (4.6) has such a solution, which holds with the coefficients defined in A, B, and C (see Boussama *et al.* 2011).

Each sample of length N from the GARCH-type processes were produced after discarding a burn-in sample of length 50 starting from an initial innovation. In the simulation, we consider samples sizes of 125, 250 and 500, which roughly match the number of trading days in a quarter, half a year, one year, and two years, respectively.

Table 4.1 displays the percentage of rejections of $\mathcal{H}_0$ using the two proposed test statistics $V_{N,K}$ and $M_{N,K}$ based on 1000 independent simulations from each DGPs for several choices of K and nominal levels of 10%, 5% and 1%. Both test statistics show reasonably good size in finite samples that improve with increasing N, in accordance with Theorem 2.1. This also suggests that the Welch-Sattherwaite style approximation for the limiting distribution of $M_{N,K}$ performs well.

Regarding the power of each test, we noticed that in general the test based on $M_{N,K}$ had greater power than the test based on $V_{N,K}$ for the examples considered in the simulation. Increasing K in general reduces the power of the tests, which is expected in these examples, since the level of serial correlation in the squared processes is decreasing at higher lags. However, this is not always the case when these test statistics are used as a diagnostic of fitted FGARCH models below, since in that case serial correlation in the squared process is not necessarily monotonically decreasing with increasing lags. Additionally, in the case of PWGARCH model, the power of $V_{N,K}$ test decays more slowly than the $M_{N,K}$ test as K increases.

4.3. **Simulation study of FGARCH goodness-of-fit tests.** We now turn to a simulation study of the proposed test statistics applied to diagnostic checking of FGARCH models as described in Section 3. In particular, we generate data from the following three DGPs: the FARCH(1), FARCH(2), and FGARCH(1,1). The specific FARCH(2) model considered is defined as

$$X_i(t) = \sigma_i(t)\varepsilon_i(t)$$

where $\varepsilon_i(t)$ is defined in (4.4) and,

$$\sigma_i^2(t) = \omega(t) + \int 12(t\cdot(1-t))(s\cdot(1-s))X_{i-1}^2(s)ds + \int 12(t\cdot(1-t))(s\cdot(1-s))X_{i-2}^2(s)ds.$$

For each simulated sample we then test for the model adequacy of the FARCH(1) model. When the data follows the FARCH(1) specification, we expect the test to reject the adequacy of the FARCH(1) model only a specified level of significance, while we expect that the adequacy of

TABLE 4.1. Empirical rejection rates of the tests for conditional heteroscedasticity using $V_{N,K}$ and $M_{N,K}$ based on 1000 independent simulations at asymptotic levels of 10%, 5%, and 1%.

DGP:		IID-BM		FARCH(1)		FGARCH(1,1)		PWGARCH(1,1)		FBEKK(1,1)	
Statistic:		$V_{N,K}$	$M_{N,K}$	$V_{N,K}$	$M_{N,K}$	$V_{N,K}$	$M_{N,K}$	$V_{N,K}$	$M_{N,K}$	$V_{N,K}$	$M_{N,K}$
K=1											
N=125	10%	0.07	0.07	0.93	0.98	0.63	0.80	0.63	0.78	0.40	0.40
	5%	0.04	0.04	0.91	0.97	0.56	0.75	0.56	0.73	0.32	0.33
	1%	0.01	0.01	0.80	0.94	0.41	0.63	0.39	0.61	0.18	0.20
N=250	10%	0.07	0.07	1.00	1.00	0.89	0.97	0.90	0.96	0.70	0.71
	5%	0.04	0.04	1.00	1.00	0.85	0.96	0.85	0.94	0.62	0.64
	1%	0.01	0.01	0.99	1.00	0.75	0.92	0.76	0.91	0.49	0.51
N=500	10%	0.10	0.09	1.00	1.00	0.99	1.00	1.00	1.00	0.95	0.95
	5%	0.05	0.05	1.00	1.00	0.99	1.00	0.99	1.00	0.92	0.92
	1%	0.01	0.01	1.00	1.00	0.97	1.00	0.97	0.99	0.85	0.86
K=5											
N=125	10%	0.07	0.08	0.81	0.92	0.67	0.89	0.68	0.90	0.59	0.60
	5%	0.04	0.05	0.75	0.89	0.63	0.88	0.60	0.86	0.53	0.55
	1%	0.01	0.02	0.60	0.83	0.52	0.81	0.50	0.79	0.41	0.44
N=250	10%	0.08	0.08	0.98	0.99	0.93	0.99	0.93	0.99	0.89	0.90
	5%	0.04	0.05	0.98	0.99	0.91	0.99	0.89	0.99	0.87	0.88
	1%	0.01	0.02	0.94	0.99	0.84	0.99	0.84	0.98	0.80	0.82
N=500	10%	0.09	0.09	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00
	5%	0.05	0.05	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00
	1%	0.01	0.02	1.00	1.00	0.99	1.00	0.99	1.00	0.98	0.99
K=10											
N=125	10%	0.06	0.06	0.76	0.86	0.60	0.86	0.66	0.88	0.56	0.57
	5%	0.03	0.03	0.68	0.82	0.55	0.82	0.59	0.85	0.49	0.50
	1%	0.01	0.01	0.53	0.75	0.43	0.75	0.49	0.79	0.38	0.40
N=250	10%	0.08	0.08	0.97	0.99	0.90	0.99	0.90	0.99	0.91	0.92
	5%	0.04	0.04	0.95	0.98	0.87	0.99	0.86	0.99	0.88	0.88
	1%	0.01	0.01	0.92	0.97	0.82	0.98	0.79	0.98	0.82	0.83
N=500	10%	0.10	0.09	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
	5%	0.05	0.06	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00
	1%	0.01	0.02	1.00	1.00	0.99	1.00	0.98	1.00	0.99	0.99
K=20											
N=125	10%	0.05	0.05	0.52	0.91	0.71	0.77	0.53	0.21	0.18	1.00
	5%	0.02	0.02	0.45	0.86	0.66	0.72	0.48	0.19	0.13	0.99
	1%	0.01	0.01	0.32	0.73	0.55	0.62	0.40	0.17	0.10	0.96
N=250	10%	0.07	0.07	0.90	0.99	0.96	0.94	0.85	0.42	0.29	1.00
	5%	0.04	0.03	0.86	0.97	0.96	0.90	0.81	0.40	0.23	1.00
	1%	0.01	0.01	0.74	0.91	0.92	0.83	0.74	0.35	0.13	1.00
N=500	10%	0.09	0.09	1.00	1.00	1.00	1.00	0.98	0.75	0.52	1.00
	5%	0.05	0.05	1.00	1.00	1.00	0.99	0.98	0.72	0.32	1.00
	1%	0.01	0.01	0.98	1.00	1.00	0.98	0.97	0.65	0.17	1.00

the FARCH(1) model is rejected at a high rate for data generated according to the FARCH(2) and FGARCH(1,1) models. To estimate these models, we set $L = 1$ in (3.1). Table 4.2 displays the rejection rates of each model using the test statistics $V_{N,K,\varepsilon}^{heuristic}$, $M_{N,K,\varepsilon}^{heuristic}$ and $M_{N,K,\varepsilon}$ for each DGP and with increasing values of N and K. Both heuristic tests are shown to have a reasonable size for the fitted residuals, although the test based on $M_{N,K,\varepsilon}^{heuristic}$ test was somewhat over-sized in large samples. Both tests perform well in detecting mis-specified models, with increasingly better performance for larger sample sizes. Similar to the results obtained in the last subsection, the $V_{N,K,\varepsilon}^{heuristic}$ test is comparably less powerful than the $M_{N,K,\varepsilon}^{heuristic}$ test. As a comparison to $M_{N,K,\varepsilon}^{heuristic}$ test, the asymptotic $M_{N,K,\varepsilon}$ test exhibits a improved size when $K = 1$ and 5 under $\mathcal{H}_0$, and slightly less power under $\mathcal{H}_A$, and this is accordance with our expectation because the asymptotic result is sharper for the latter statistic. The tests become in accordance with the corrected size and slightly over-sized when $K = 10$ and 20 and correspondingly more powerful under $\mathcal{H}_A$, we attribute this to the increased error from the number of performed Monte Carlo integration.

Another observation worthy of a remark is that the rejection rates of the adequacy of the FARCH(1) model tend to be low for all DGP when $K = 1$. This is because fitting a FARCH(1) model tends to remove serial correlation from the squared process at lag one. Hence it is advisable when using this test for the purpose of model diagnostic checking to incorporate several lags beyond the order of the applied model.

One avenue that we investigate further is whether or not the size inflation of each test could be explained by the sampling variability of the estimates of the principal components of the squared process. In order to evaluate this, we perform the same simulation as described above, but with the first principal component $\hat{\phi}_1(t)$ being replaced by the “oracle" basis function

$$\phi_1(t) = t(1-t)/\left\|t(1-t)\right\|.$$

Using this function in the basis to reduce the dimension of the operators to be estimated is ideal since for the processes that we consider the operators defining them are rank one with a range spanned by ϕ_1. The rejection rates of the adequacy of each model with this modification to the tests are displayed in Table 4.3, which shows that both the size and the power of the test in general are somewhat improved for all tests. This simulation result suggests that we can improve the estimation of the FGARCH models by changing the basis used for dimension reduction, although it is in general not clear how to improve upon the FPCA method; doing so is beyond the scope of the current paper.

Table 4.2. Rejection rates from 1000 independent simulations of the model adequacy of the FARCH(1) model when applied to FARCH(1), FARCH(2), and FGARCH(1,1) data.

DGP:		FARCH(1)			FARCH(2)			FGARCH(1,1)		
Statistics:		$V_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}$	$V_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}$	$V_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}$
					K=1					
	10%	0.06	0.06	0.07	0.07	0.07	0.08	0.07	0.07	0.09
N=125	5%	0.03	0.03	0.03	0.04	0.05	0.06	0.04	0.04	0.06
	1%	0.01	0.01	0.01	0.02	0.02	0.02	0.01	0.02	0.02
	10%	0.10	0.09	0.10	0.09	0.08	0.11	0.11	0.11	0.12
N=250	5%	0.05	0.06	0.05	0.05	0.07	0.08	0.05	0.07	0.09
	1%	0.02	0.02	0.01	0.02	0.03	0.04	0.02	0.04	0.04
	10%	0.15	0.17	0.14	0.16	0.15	0.14	0.17	0.16	0.16
N=500	5%	0.10	0.11	0.08	0.11	0.10	0.09	0.11	0.11	0.10
	1%	0.03	0.05	0.03	0.05	0.06	0.04	0.05	0.05	0.03
					K=5					
	10%	0.07	0.07	0.07	0.64	0.67	0.60	0.39	0.44	0.44
N=125	5%	0.03	0.04	0.04	0.55	0.60	0.50	0.30	0.35	0.34
	1%	0.01	0.01	0.01	0.42	0.48	0.32	0.19	0.24	0.22
	10%	0.08	0.08	0.07	0.89	0.90	0.87	0.71	0.73	0.74
N=250	5%	0.04	0.05	0.04	0.84	0.86	0.82	0.64	0.67	0.65
	1%	0.02	0.03	0.01	0.73	0.77	0.68	0.48	0.55	0.49
	10%	0.13	0.12	0.12	0.99	1.00	0.98	0.92	0.93	0.94
N=500	5%	0.08	0.08	0.06	0.99	0.99	0.97	0.89	0.90	0.89
	1%	0.02	0.03	0.01	0.97	0.98	0.93	0.80	0.84	0.79
					K=10					
	10%	0.05	0.06	0.09	0.51	0.52	0.53	0.33	0.37	0.40
N=125	5%	0.03	0.03	0.05	0.41	0.45	0.44	0.27	0.31	0.28
	1%	0.01	0.01	0.02	0.29	0.34	0.30	0.18	0.22	0.17
	10%	0.08	0.08	0.11	0.82	0.84	0.81	0.63	0.65	0.68
N=250	5%	0.04	0.06	0.06	0.76	0.79	0.72	0.54	0.57	0.57
	1%	0.01	0.02	0.01	0.63	0.68	0.58	0.40	0.46	0.39
	10%	0.11	0.12	0.17	0.98	0.98	0.98	0.89	0.89	0.91
N=500	5%	0.06	0.07	0.08	0.98	0.98	0.97	0.87	0.88	0.87
	1%	0.02	0.03	0.03	0.95	0.96	0.94	0.72	0.77	0.76
					K=20					
	10%	0.03	0.03	0.12	0.41	0.39	0.42	0.23	0.24	0.31
N=125	5%	0.02	0.02	0.06	0.35	0.33	0.34	0.18	0.18	0.21
	1%	0.01	0.01	0.02	0.22	0.23	0.20	0.11	0.12	0.12
	10%	0.07	0.06	0.15	0.75	0.73	0.79	0.49	0.51	0.57
N=250	5%	0.04	0.03	0.08	0.66	0.66	0.70	0.43	0.44	0.45
	1%	0.01	0.01	0.02	0.51	0.54	0.53	0.30	0.33	0.32
	10%	0.10	0.10	0.16	0.96	0.96	0.95	0.81	0.82	0.85
N=500	5%	0.05	0.06	0.08	0.94	0.94	0.93	0.76	0.78	0.75
	1%	0.02	0.02	0.02	0.88	0.89	0.85	0.63	0.66	0.61

5. Application to dense intra-day asset price data

A natural example of functional financial time series data are those derived from densely recorded asset price data, such as intraday stock price data. Recently there has been a great deal of research focusing on analyzing the information contained in the curves constructed from such data. Price curves associated with popular companies are routinely displayed for public review. The objectives of this section are to 1) test whether functional financial time series derived

Table 4.3. Rejection rates from 1000 independent simulations of the model adequacy of the FARCH(1) model when applied to FARCH(1), FARCH(2), and FGARCH(1,1) data when the first basis function used for dimension reduction is ϕ_1.

DGP:		FARCH(1)			FARCH(2)			FGARCH(1,1)		
Statistics:		$V_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}$	$V_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}$	$V_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}^{heuristic}$	$M_{N,K,\varepsilon}$
						K=1				
	10%	0.06	0.07	0.07	0.05	0.06	0.06	0.06	0.06	0.08
N=125	5%	0.03	0.04	0.03	0.03	0.03	0.03	0.03	0.04	0.04
	1%	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.01
	10%	0.10	0.11	0.10	0.08	0.08	0.09	0.11	0.10	0.10
N=250	5%	0.06	0.07	0.04	0.05	0.05	0.05	0.06	0.07	0.05
	1%	0.02	0.03	0.01	0.02	0.02	0.02	0.02	0.03	0.01
	10%	0.16	0.16	0.13	0.15	0.13	0.15	0.15	0.15	0.14
N=500	5%	0.10	0.11	0.06	0.10	0.10	0.09	0.10	0.10	0.08
	1%	0.04	0.05	0.02	0.05	0.06	0.03	0.04	0.05	0.03
						K=5				
	10%	0.06	0.08	0.09	0.59	0.62	0.65	0.42	0.47	0.44
N=125	5%	0.04	0.04	0.04	0.52	0.56	0.56	0.33	0.37	0.33
	1%	0.01	0.02	0.01	0.36	0.42	0.37	0.19	0.25	0.20
	10%	0.11	0.12	0.10	0.90	0.91	0.87	0.68	0.72	0.71
N=250	5%	0.07	0.08	0.05	0.86	0.87	0.81	0.59	0.65	0.62
	1%	0.02	0.03	0.01	0.75	0.78	0.65	0.45	0.50	0.44
	10%	0.14	0.14	0.13	0.99	1.00	0.99	0.92	0.93	0.93
N=500	5%	0.08	0.09	0.07	0.99	0.99	0.98	0.87	0.89	0.87
	1%	0.02	0.04	0.01	0.96	0.97	0.95	0.78	0.78	0.77
						K=10				
	10%	0.06	0.06	0.09	0.51	0.53	0.55	0.35	0.38	0.41
N=125	5%	0.04	0.04	0.04	0.43	0.45	0.45	0.27	0.32	0.31
	1%	0.01	0.01	0.01	0.29	0.35	0.30	0.17	0.22	0.18
	10%	0.08	0.08	0.11	0.83	0.84	0.82	0.66	0.69	0.70
N=250	5%	0.04	0.06	0.06	0.78	0.80	0.74	0.56	0.61	0.59
	1%	0.01	0.01	0.01	0.66	0.71	0.59	0.41	0.48	0.41
	10%	0.11	0.13	0.15	0.99	0.99	0.98	0.89	0.90	0.91
N=500	5%	0.07	0.08	0.07	0.98	0.98	0.97	0.85	0.87	0.87
	1%	0.03	0.04	0.02	0.95	0.96	0.92	0.73	0.79	0.75
						K=20				
	10%	0.06	0.05	0.11	0.42	0.39	0.43	0.26	0.26	0.31
N=125	5%	0.03	0.03	0.06	0.34	0.33	0.35	0.19	0.20	0.24
	1%	0.01	0.01	0.02	0.23	0.24	0.19	0.13	0.14	0.12
	10%	0.07	0.06	0.14	0.78	0.77	0.79	0.51	0.52	0.57
N=250	5%	0.04	0.04	0.07	0.71	0.72	0.72	0.44	0.46	0.46
	1%	0.02	0.01	0.02	0.58	0.62	0.55	0.32	0.36	0.31
	10%	0.10	0.09	0.16	0.97	0.97	0.96	0.82	0.83	0.86
N=500	5%	0.06	0.06	0.08	0.94	0.94	0.93	0.75	0.76	0.75
	1%	0.02	0.02	0.02	0.88	0.89	0.86	0.63	0.65	0.62

from the dense intraday price data exhibit conditional heteroscedasticity, and 2) evaluate the adequacy of FGARCH models for such series.

The specific data that we consider consists of 5 minute resolution closing prices of Standard & Poor's 500 market index, so that there are $J = 78$ observations of the closing price each day. For the purpose of applying a Monte Carlo integration to the asymptotic diagnostic test $M_{N,K,\varepsilon}$, we employ a sparse grid of $J^* = 39$ out of the 78 points. Then, we let $P_i(t)$ denote the price of either asset on day i at intraday time t, where t is normalized to the unit interval.

We consider time series of curves from these data of length $N = 502$ taken from the dates between 31/December/2015 to 02/January/2018. There are several ways to define curves that are approximately stationary based on the raw price curves $P_i(t)$. We consider the following three cases:

(1) Overnight cumulative intra-day log returns (OCIDRs)

$$X_i(t) = \log P_i(t) - \log P_{i-1}(1)$$

(2) Cumulative intra-day log returns (CIDRs)

$$X_i(t) = \log P_i(t) - \log P_i(0)$$

(3) Intra-day log returns (IDRs)

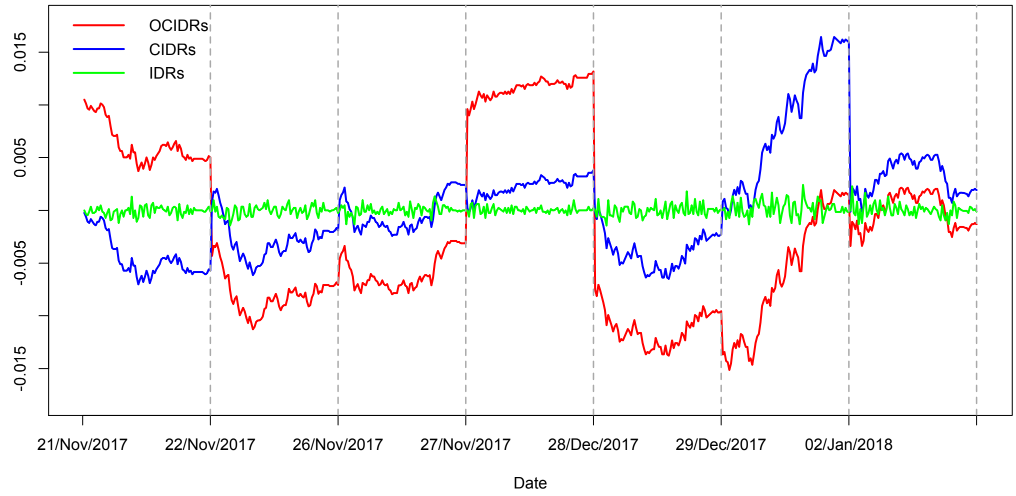
$$X_i(t) = \log P_i(t) - \log P_i(t-w)$$

The later two functions have been studied in the literature, and the first function measures the trajectory of cumulative price changes between the current intra-day price and the closing price from the previous day. A similar overnight return has been used in Koopman *et al.* (2005). To obtain the IDRs, we use $w = 1$ to produce the 5-min intra-day log returns. Figure 5.1 shows these three types of intra-day curves across seven days. The stationarity of all three return curves is examined by using the stationary tests proposed by Horváth *et al.* (2014). The results suggest that all intra-day return series are stationary.

We begin by testing for functional conditional heteroscedasticity for each curve type. The results of these tests are given in Table 5.1, which suggest that each sample of curves exhibit strong conditional heteroscedasticity.

A natural next step is to posit and evaluate models to capture this conditional heteroscedasticity. For this we consider two models: standard scalar GARCH models and FGARCH models. The motivation for considering standard scalar GARCH models for this purpose is that we might at first expect that the volatility in each of these curves can be adequately accounted for by scaling each curve by the conditional standard deviation through the fitting of a GARCH model to the end-of-day returns, *i.e.* a large magnitude of the return on the previous day spells high volatility for the entire intraday price on the following day. We compute the daily log returns

FIGURE 5.1. Seven days of Intra-day return curves from S&P 500



as $x_i = \log(P_i(1)) - \log(P_{i-1}(1))$, $1 \le i \le 500$, to which we fit a scalar GARCH(p,q) model by using a quasi maximum likelihood estimation approach. The orders $\{p, q\}$ are selected as the minimum orders for which the estimated residuals $\hat{\varepsilon}_i = x_i/\hat{\sigma}_i$ are plausibly a strong white noise as measured by the Li-Mak test; see Li and Mak (1994), resulting in the selection of a GARCH(1,1) model, as shown in Panel A in Table 5.2.

We then apply the proposed tests for conditional heteroscedasticity to the fitted residuals functions of intra-day returns

$$\tilde{\varepsilon}_i(t) = X_i(t)/\hat{\sigma}_i.$$

The results of these tests are given in Panel B in Table 5.2, which show that these curves still exhibit a substantial amount of conditional heteroscedasticity.

Next, we consider the FARCH(1) and FGARCH(1,1) models for these curves. We fit each model with $L = 1$ in (3.1) in order to be consistent with the simulation section, and evaluate the adequacy of each model as proposed above. Figure 5.2 shows plots of $w(t)$ and wire-frame plots of the kernels $\alpha(t, s)$ and $\beta(t, s)$ for the FGARCH(1,1) model for each type of intra-day return curve. We then estimate $\hat{\sigma}_i(t)$ recursively with the initial values of $\hat{w}(t)$, and the de-volatized intra-day return $\hat{\varepsilon}_i(t)$ is fitted per Equation (3.3). Figure 5.3 exhibits the de-volatized intra-day returns over seven days by using the FGARCH(1,1) model.

Table 5.3 reports the p-values from the diagnostic checks of the FGARCH(1,1) and FARCH(1) models applied to de-volatized intra-day returns. All of the three diagnostic tests show broadly consistent results at specified significance levels. The FARCH(1) model is generally deemed to be inadequate for each curve type, although this model performs as we expected to adequately model conditional heteroscedasticity at lag 1. By contrast, the P values in Panel B of Table 5.3 suggest that the FGARCH(1,1) model is generally acceptable for modelling the conditional heteroscedasticity of all three curve types. In conjunction with the above results showing that these curves cannot be adequately de-volatized simply by scaling with the conditional standard deviation estimates from GARCH models for the scalar returns, we draw the following tentative conclusions from this analysis: 1) the magnitude of the return cannot fully explain the volatility of intraday prices observed on subsequent days; instead we should consider the entire path of the price curve on previous days in order to adequately model future intra-day conditional heteroscedasticity, and 2) the FGARCH class of models seems to be effective for modeling intra-day conditional heteroscedasticity.

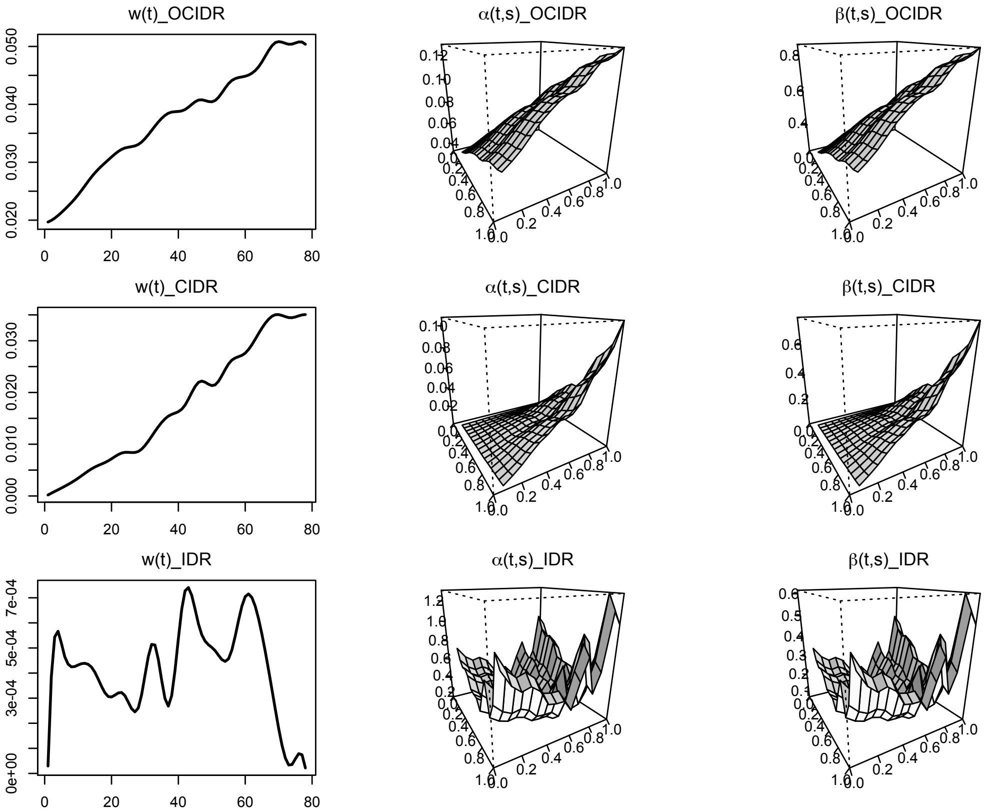
TABLE 5.1. Heteroscedasticity tests on the intra-day returns of S&P 500

		K=1		K=5		K=10		K=20	
		Stats	P value	Stats	P value	Stats	P value	Stats	P value
OCIDRs	$V_{N,K}$	6.94	0.01	52.69	0.00	73.70	0.00	76.20	0.00
	$M_{N,K}$	1.19	0.00	8.44	0.00	12.08	0.00	12.82	0.00
CIDRs	$V_{N,K}$	8.73	0.00	36.11	0.00	37.76	0.00	49.35	0.00
	$M_{N,K}$	0.06	0.00	0.26	0.00	0.29	0.00	0.42	0.00
IDRs	$V_{N,K}$	189.87	0.00	437.99	0.00	461.05	0.00	481.15	0.00
	$M_{N,K}$	4.73e-08	0.00	1.25e-07	0.00	1.55e-07	0.00	2.30e-07	0.00

TABLE 5.2. Heteroscedasticity tests of de-volatized return curves $\tilde{\varepsilon}_i(t)$ using a GARCH(p,q) model

Panel A: Li-Mak Test on $\hat{\varepsilon}_i$								
	K=1		K=5		K=10		K=20	
Model	Stats	P value	Stats	P value	Stats	P value	Stats	P value
GARCH(1,1)	0.07	0.79	2.50	0.77	10.91	0.36	26.94	0.13
Panel B: $V_{N,K}$ and $M_{N,K}$ Test on $\tilde{\varepsilon}_i(t)$								
OCIDRs								
$V_{N,K}$	4.82	0.03	38.26	0.00	72.89	0.00	75.18	0.00
$M_{N,K}$	2.5e+17	0.01	1.7e+18	0.00	3.5e+18	0.00	4.3e+18	0.00
CIDRs								
$V_{N,K}$	5.99	0.01	25.86	0.00	28.78	0.00	38.28	0.01
$M_{N,K}$	1.51e+16	0.01	7.59e+16	0.00	1.01e+17	0.00	1.71e+17	0.00
IDRs								
$V_{N,K}$	165.27	0.00	348.15	0.00	377.97	0.00	410.44	0.00
$M_{N,K}$	9.09e+09	0.00	2.70e+10	0.00	3.90e+10	0.00	6.21e+10	0.00

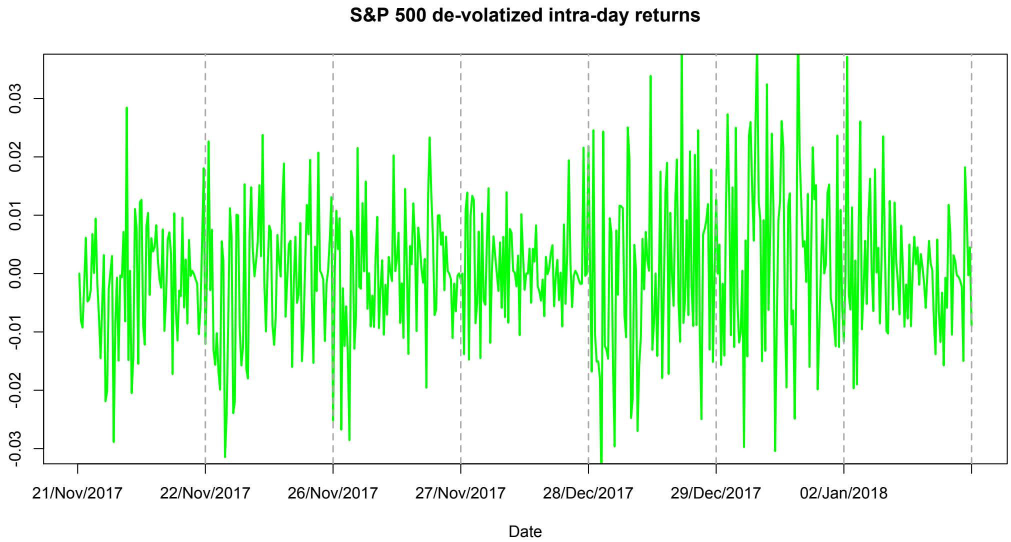
FIGURE 5.2. Plots of the estimated kernels for the FGARCH(1,1) model for the S&P 500 intra-day return curves



6. CONCLUSION

We proposed two portmanteau-type conditional heteroscedasticity tests for functional time series. By applying the test statistics to model residuals from the fitted functional GARCH models, our tests also provide two heuristic and one asymptotically valid goodness-of-fit test for such models. Simulation results presented in this paper show that both tests have good size and power to detect conditional heteroscedasticity in functional financial time series and assess the goodness-of-fit of the FGARCH models in finite samples. In an application to the dense intraday price data, we investigated the conditional heteroscedasticity of three types of the intra-day return curves, including the overnight cumulative intra-day returns, the cumulative intra-day returns and the intra-day log returns from two assets. Our results suggested that these curves exhibit substantial evidence of conditional heteroscedasticity that cannot be accounted for simply by rescaling the curves by using measurements of the conditional standard deviation based on the magnitude of the scalar returns. However, the functional conditional volatility models often

FIGURE 5.3. Plots of de-volatized S&P 500 intra-day return curves based on an FGARCH(1,1) Model



appeared to be adequate for modeling this observed functional conditional heteroscedasticity in financial data.

APPENDIX A. PROOFS OF RESULTS IN SECTIONS 2

Proofs of Theorem 2.1. First we show (2.5). Under $\mathcal{H}_0$ and Assumption 2.1 the random variables $Y_{1,i} = \|X_i\|^2$ are independent and identically distributed, and satisfy $EY_{1,i}^4 < \infty$. (2.5) now follows Theorem 7.2.1 and problem 2.19 of Brockwell and Davis (1991).

In order to show (2.6), we recall some notation and the statement of Lemma 5 in Kokoszka *et al.* (2017). Let K be a positive integer as in the definition of $M_{N,K}$. Consider the space $\mathcal{G}_1$

TABLE 5.3. Diagnostic tests of FGARCH(1,1) and FARCH(1) models applied to the S&P 500 return curves.

		Panel A: FARCH(1)							
		Stats	P value	Stats	P value	Stats	P value	Stats	P value
OCIDRs	$V_{N,K,\varepsilon}^{heuristic}$	0.26	0.61	33.94	0.00	50.07	0.00	52.53	0.00
	$M_{N,K,\varepsilon}^{heuristic}$	5.86	0.52	216.08	0.00	324.45	0.00	358.37	0.00
	$M_{N,K,\varepsilon}$	5.86	0.43	216.08	0.00	324.45	0.00	358.37	0.00
CIDRs	$V_{N,K,\varepsilon}^{heuristic}$	2.40	0.12	33.11	0.00	36.59	0.00	52.32	0.00
	$M_{N,K,\varepsilon}^{heuristic}$	66.07	0.05	586.09	0.00	737.63	0.00	1264.67	0.00
	$M_{N,K,\varepsilon}$	66.07	0.14	586.09	0.00	737.63	0.00	1264.67	0.00
IDRs	$V_{N,K,\varepsilon}^{heuristic}$	1.16	0.28	20.84	0.00	26.02	0.00	45.91	0.00
	$M_{N,K,\varepsilon}^{heuristic}$	120.50	1.00	2942.47	0.34	4299.53	0.97	9407.58	0.96
	$M_{N,K,\varepsilon}$	120.50	0.98	2942.47	0.00	4299.53	0.00	9407.58	0.00
		Panel B: FGARCH(1,1)							
		K=1		K=5		K=10		K=20	
		Stats	P value	Stats	P value	Stats	P value	Stats	P value
OCIDRs	$V_{N,K,\varepsilon}^{heuristic}$	0.11	0.74	2.15	0.83	5.80	0.83	8.88	0.98
	$M_{N,K,\varepsilon}^{heuristic}$	5.11	0.59	27.36	0.84	65.30	0.85	106.47	0.99
	$M_{N,K,\varepsilon}$	5.11	1.00	27.36	1.00	65.30	1.00	106.47	1.00
CIDRs	$V_{N,K,\varepsilon}^{heuristic}$	0.17	0.68	4.00	0.55	8.55	0.58	18.40	0.56
	$M_{N,K,\varepsilon}^{heuristic}$	25.01	0.87	232.20	0.59	469.33	0.61	923.36	0.72
	$M_{N,K,\varepsilon}$	25.01	1.00	232.20	1.00	469.33	1.00	923.36	1.00
IDRs	$V_{N,K,\varepsilon}^{heuristic}$	3.95	0.05	11.54	0.04	16.77	0.08	33.25	0.03
	$M_{N,K,\varepsilon}^{heuristic}$	301.90	1.00	3912.94	1.00	7170.58	1.00	17859.05	1.00
	$M_{N,K,\varepsilon}$	301.90	0.99	3912.94	0.63	7170.58	0.91	17859.05	0.16

of functions $f : [0,1]^2 \to \mathbb{R}^K$, mapping the unit square to the space of K-dimensional column vectors with real entries, satisfying

$$\iint \{f(t,s)\}^\top f(t,s)dtds < \infty.$$

This space is a separable Hilbert space when equipped with the inner product

$$\langle f, g\rangle_{\mathcal{G},1} = \iint \{f(t,s)\}^\top g(t,s)dtds.$$

Let $\|\cdot\|_{\mathcal{G},1}$ denote the norm induced by this inner product. Let $\langle\cdot,\cdot\rangle_{\mathbb{F}}$ denote the matrix Frobenius inner product, and let $\|\cdot\|_{\mathbb{F}}$ denote the corresponding norm; see Chapter 5 of Meyer (2000). Further let $\mathcal{G}_2$ denote the space of functions $f : [0,1]^4 \to \mathbb{R}^{K\times K}$, equipped with the inner product

$$\langle f, g\rangle_{\mathcal{G},2} = \iiiint \langle f(t,s,u,v), g(t,s,u,v)\rangle_{\mathbb{F}} dtdsdudv.$$

for which $\langle f, f\rangle_{\mathcal{G},2} < \infty$. $\mathcal{G}_2$ is also a separable Hilbert space when equipped with this inner product.

Let $\psi_K : [0,1]^4 \to \mathbb{R}^{K\times K}$ be a matrix valued kernel where the $1 \le i,j \le K$ component is denoted by $\psi_{K,i,j}(t,s,u,v)$. We then define ψ_K by

$$\psi_{K,i,j}(t,s,u,v) = \begin{cases} \text{cov}(X_0^2(t), X_0^2(u))\text{cov}(X_0^2(s), X_0^2(v)), & 1 \le i = j \le K. \\ 0 \;\; 1 \le i \ne j \le K. \end{cases} \tag{A.1}$$

The kernel ψ_K defines a linear operator $\Psi_K : \mathcal{G}_1 \to \mathcal{G}_1$ by

$$\Psi_K(f)(t,s) = \iint \psi_K(t,s,u,v) f(u,v)dudv, \tag{A.2}$$

where the integration is carried out coordinate-wise. Following the preamble to the proof of Lemma 5 of Kokoszka *et al.* (2017), it follows that the operator Ψ_K is compact, symmetric, and positive definite. Due to these three properties, we have by the spectral theorem for positive definite, self-adjoint, compact operators, *e.g.* Chapter 6.2 of Riesz and Nagy (1990), that Ψ_K defines a nonnegative and decreasing sequence of eigenvalues and a corresponding orthonormal basis of eigenfunctions $\varphi_{i,K}(t,s)$, $1 \le i < \infty$, satisfying

$$\Psi_K(\varphi_{i,K})(t,s) = \xi_{i,K}\varphi_{i,K}(t,s), \quad \text{with} \sum_{i=1}^{\infty} \xi_{i,K} < \infty. \tag{A.3}$$

With this notation, we now define

$$\hat{\Gamma}_{N,K}(t,s) = \sqrt{N}\left\{\hat{\gamma}_1(t,s), \ldots, \hat{\gamma}_K(t,s)\right\}^{\top} \in \mathcal{G}_1.$$

Under $\mathcal{H}_0$ and Assumption 2.1, the sequence $\{X_i^2(t)\}$ satisfies the conditions of Lemma 5 of Kokoszka *et al.* (2017), which implies that $\hat{\Gamma}_{N,K}(t,s) \overset{\mathcal{D}(\mathcal{G}_1)}{\to} \Gamma_K(t,s)$, where $\Gamma_K(t,s)$ is a Gaussian process with covariance operator Ψ_K, and $\overset{\mathcal{D}(\mathcal{G}_1)}{\to}$ denotes weak convergence in $\mathcal{G}_1$. It now follows from the Karhunen-Loéve representation and continuous mapping theorem that

$$M_{N,K} = \|\hat{\Gamma}_{N,K}\|^2 \overset{\mathcal{D}}{\to} \|\Gamma_K\|^2 \overset{\mathcal{D}}{=} \sum_{i=1}^{\infty} \xi_{i,K}\chi_1^2(i).$$

A simple calculation based on (A.1) shows that the eigenvalues of Ψ_K are products of the eigenvalues defined by (2.4), $\{\lambda_i\lambda_j,\ 1 \le i, j < \infty\}$, with each eigenvalue having multiplicity K, giving the form of the limit distribution in (2.6).

$\square$

Justification of (4.2). Using proposition 5.10.16 of Bogachev (1998), we have that

$$E(\|\Gamma_K\|^2_{\mathcal{H},1}) = \mathrm{tr}(\Psi_K) = K\left(\int \mathrm{cov}(X_0^2(t), X_0^2(t))dt\right)^2,$$

and

$$\mathrm{var}(\|\Gamma_K\|^2_{\mathcal{H},1}) = 2\,\mathrm{tr}(\Psi_K^2) = 2K\left(\iint \mathrm{cov}(X_0^2(t), X_0^2(s))dtds\right)^2.$$

Proof of Theorem 2.2. We only show (2.7) as (2.8) follows similarly from it. Let $C_h(t,s) = \mathrm{cov}(X_i^2(t), X_{i+h}^2(s)) \neq 0$. It follows from the assumed L^8-m-approximability of X_i that X_i^2 is L^4-m-approximable, from which we can show that

$$\|\hat{\gamma}_h(t,s) - C_h(t,s)\| = O_P(1/\sqrt{N}).$$

Now $M_{N,K} \ge N\|\hat{\gamma}_h(t,s)\|^2$, and $\|\hat{\gamma}_h(t,s)\|^2 = \|\hat{\gamma}_h - C_h\|^2 + 2\langle\hat{\gamma}_h - C_h, C_h\rangle + \|C_h\|^2$. It follows that $N[\|\hat{\gamma}_h - C_h\|^2 + 2\langle\hat{\gamma}_h - C_h, C_h\rangle] = O_P(\sqrt{N})$, and $N\|C_h\|^2$ diverges to positive infinity at rate N, yielding the desired result. $\square$

Proof of Theorem 2.3. Again we only prove (2.10) as (2.9) follows from it by a similar argument. By squaring both sides of of (2.1) and iterating (2.2), we obtain that

$$X_i^2(t) = \omega_\alpha(t) + \sum_{\ell=0}^{\infty} \boldsymbol{\alpha}^{(\ell)}(V_{i-\ell})(t),$$

where the series on the right hand side of the above equation converges in $L^2[0,1]$ with probability one, and

$$\omega_\alpha(t) = \sum_{\ell=0}^{\infty} \boldsymbol{\alpha}^{(\ell)}(\omega)(t).$$

Therefore, $X_i^2(t)$ is a linear process in $L^2[0,1]$ with mean $\omega_\alpha(t)$ generated by the weak functional white noise innovations V_i as defined in Bosq (2000). It now follows from Assumption 2.1 and the ergodic theorem that

$$\left\| \hat{\gamma}_h(t,s) - \sum_{j=0}^{\infty} E\boldsymbol{\alpha}^{(j)}(V_j)(t)\boldsymbol{\alpha}^{(j+h)}(V_j)(s) \right\| = o_P(1).$$

It follows from this and the reverse triangle inequality that

$$\frac{M_{N,K}}{N} = \sum_{h=1}^{K} \|\hat{\gamma}_h\|^2 \xrightarrow{p} \sum_{h=1}^{K} \left\| \sum_{j=0}^{\infty} E\boldsymbol{\alpha}^{(j)}(V_j)(t)\boldsymbol{\alpha}^{(j+h)}(V_j)(s) \right\|^2,$$

as desired.

□

Appendix B. Proof of Theorem 3.1 and estimation of parameters/kernels in Section 3.1

We first develop some notation and detail the assumptions that we use to establish Theorem 3.1. Recall from equation (3.2) that the FGARCH equations along with (3.1) imply that

$$\mathfrak{s}_i^2 = D + A\mathbf{x}_{i-1}^2 + B\mathfrak{s}_{i-1}^2$$

where $\mathbf{x}_i^2 = [\langle X_i^2(t), \phi_1(t)\rangle, \ldots, \langle X_i^2(t), \phi_L(t)\rangle]^\top$, $\mathfrak{s}_i^2 = [\langle \sigma_i^2(t), \phi_1(t)\rangle, \ldots, \langle \sigma_i^2(t), \phi_L(t)\rangle]^\top$, the coefficient vector $D = [d_1, \ldots, d_L]^\top \in \mathbb{R}^L$, and the coefficient matrices A and B are $\mathbb{R}^{L\times L}$ with (j, j') entries by $a_{j,j'}$ and $b_{j,j'}$, respectively. Let $\Gamma_0(t,s) = \alpha(t,s)\varepsilon_0^2(s) + \beta(t,s)$. We make the following assumptions:

Assumption B.1. $E\left\|\int \Gamma_0(\cdot, s)ds\right\|_\infty^2 < 1$, *and* $\omega \in C[0,1]$.

Assumption B.2. Q_0 *is nonsingular.*

Assumption B.3. $\mathbf{x}_0^2$ *is not measurable with respect to* $\mathcal{F}_0$.

Assumption B.4. $\inf_{\theta\in\Theta} |det(A)| > 0$ *and* $\sup_{\theta\in\Theta} \|B\|_{op} < 1$, *where* $\|\cdot\|_{op}$ *is the matrix operator norm of* B.

Assumption B.5. $E\|\varepsilon_0^4\|_\infty < \infty$

Assumption B.6. *There exists a constant* δ *so that* $\inf_{\theta\in\Theta} \inf_{t\in[0,1]} \omega(t) \geq \delta > 0$.

Assumptions B.1–B.4 come directly from Aue et al. (2017), and imply both that there exists a strictly stationary and causal solution to the FGARCH equations in $C[0,1]$, and that $\hat{\theta}_N$ is a strongly consistent estimator of θ_0 that also satisfies the central limit theorem. Assumption B.5 appears in Cerovecki et al. (2019), and is a somewhat stronger assumption than that of Theorem 3.2 of Aue et al (2018). It is used in the proofs below mainly to establish uniform integrability of terms of the form $\|X_i/\sigma_i\|_\infty$. Assumption B.6 is implied by the conditions of Cerovecki et al. (2019) that the functions ϕ_i are strictly positive and that $D \in \Theta_D \subset (0,\infty)^L$, where Θ_D is compact, but also may hold under more general conditions.

Theorem B.1 (Precise statement of Theorem 3.1:)**.** *Let* $\Gamma_{N,K}^{(\varepsilon,\theta)} = (\sqrt{N}\hat{\gamma}_{\varepsilon,1}, ..., \sqrt{N}\hat{\gamma}_{\varepsilon,K})^\top \in \mathcal{G}_1$. *Then under Assumption B.1–B.6,*

$$\Gamma_{N,K}^{(\varepsilon,\theta)} \overset{\mathcal{D}(\mathcal{G}_1)}{\to} \Gamma_{\varepsilon,\theta},$$

where $\Gamma_{\varepsilon,\theta}$ *is a mean zero Gaussian process in* $\mathcal{G}_1$ *with covariance operators* $\Psi_K^{(\varepsilon,\theta)}$ *defined by*

$$\Psi_K^{(\varepsilon,\theta)}(f)(t,s) = \iint \psi_K^{(\varepsilon,\theta)}(t,s,u,v) f(u,v)dudv, \tag{B.1}$$

where $\psi_K^{(\varepsilon,\theta)}(t,s,u,v)$ *is a matrix valued kernel defined by* (3.8)*. In addition,*

$$M_{N,K,\varepsilon} \overset{\mathcal{D}}{\to} \sum_{i=1}^{\infty} \xi_{i,K}\chi_1^2(i),$$

where $\xi_{i,K}$ $i \geq 1$ *are the eigenvalues of* $\Psi_K^{(\varepsilon,\theta)}$.

Before proving this result, we introduce further notation. We write $\sigma_i^2(t,\theta)$ to indicate the dependence of $\sigma_i^2(t)$ on the vector of parameters θ, and similarly write $\mathfrak{s}_i^2(\theta) = [\langle\sigma_i^2(t,\theta),\phi_1(t)\rangle, \ldots, \langle\sigma_i^2(t,\theta),\phi_L(t)\rangle]^\top$. It follows that with $\mathbf{\Phi}(t) = (\phi_1(t), ..., \phi_L(t))^\top$,

$$\sigma_i^2(t,\theta) = \mathfrak{s}_i^2(\theta)^\top \mathbf{\Phi}(t).$$

Iterating (3.2), we see using Assumption (B.4) that

$$\sigma_i^2(t,\theta) = \left(\sum_{\ell=0}^{\infty} B^\ell \xi_{i-\ell}\right)^\top \mathbf{\Phi}(t), \text{ where } \xi_{i-\ell} = D + A\mathbf{x}_{i-1-\ell}^2. \tag{B.2}$$

We define

$$\tilde{\sigma}_i^2(t,\theta) = \left(\sum_{\ell=0}^{i-1} B^\ell \xi_\ell\right)^\top \mathbf{\Phi}(t), \tag{B.3}$$

which allows us to define

$$\hat{\sigma}_i^2(t) = \tilde{\sigma}_i^2(t, \hat{\theta}_N). \tag{B.4}$$

In addition to $\hat{\gamma}_{\varepsilon,h}$ defined in (3.4), we also define

$$\tilde{\gamma}_{\varepsilon,h}(t,s,\theta) = \frac{1}{N}\sum_{i=1}^{N-h}\left(\frac{X_i^2(t)}{\tilde{\sigma}_i^2(t,\theta)} - 1\right)\left(\frac{X_{i+h}^2(s)}{\tilde{\sigma}_{i+h}^2(s,\theta)} - 1\right). \tag{B.5}$$

and

$$\gamma^\star_{\varepsilon,h}(t,s,\theta) = \frac{1}{N}\sum_{i=1}^{N-h}\left(\frac{X_i^2(t)}{\sigma_i^2(t,\theta)} - 1\right)\left(\frac{X_{i+h}^2(s)}{\sigma_{i+h}^2(s,\theta)} - 1\right), \tag{B.6}$$

so that $\hat{\gamma}_{\varepsilon,h}(t,s) = \tilde{\gamma}_{\varepsilon,h}(t,s,\hat{\theta}_N)$. Below we let $\theta^{(j)}$ denote the j'th coordinate of θ.

Lemma B.1. *Under Assumptions B.1–B.6, for all h such that $1 \le h \le K$,*

$$\sup_{\theta\in\Theta} \sqrt{N}\|\tilde{\gamma}_{\varepsilon,h}(\cdot,\cdot,\theta) - \gamma^\star_{\varepsilon,h}(\cdot,\cdot,\theta)\| = o_P(1), \tag{B.7}$$

and

$$\max_{j\in\{1,\dots,L+2L^2\}} \sup_{\theta\in\Theta}\left\|\frac{\partial\tilde{\gamma}_{\varepsilon,h}(\cdot,\cdot,\theta)}{\partial\theta^{(j)}} - \frac{\partial\gamma^\star_{\varepsilon,h}(\cdot,\cdot,\theta)}{\partial\theta^{(j)}}\right\| = o_P(1). \tag{B.8}$$

Proof. It follows from equation 2.4 of Aue et al. (2017) that there exists a constant $c_1 > 0$ so that almost surely

$$\sup_{\theta\in\Theta} \|\sigma_i^2(\cdot,\theta) - \tilde{\sigma}^2(\cdot,\theta)\|_\infty \leq c_1\rho^i, \quad \text{and} \quad \sup_{\theta\in\Theta} \|\sigma_i^2(\cdot,\theta) - \tilde{\sigma}^2(\cdot,\theta)\| \leq c_1\rho^i, \tag{B.9}$$

for some $0 < \rho < 1$. We then have by adding and subtracting

$$\left(\frac{X_i^2(t)}{\tilde{\sigma}_i^2(t,\theta)} - 1\right)\left(\frac{X_{i+h}^2(s)}{\sigma_{i+h}^2(s,\theta)} - 1\right)$$

in the summands defining the difference $\tilde{\gamma}_{\varepsilon,h} - \gamma_{\varepsilon,h}^\star$ that

$$\tilde{\gamma}_{\varepsilon,h}(t,s,\theta) - \gamma_{\varepsilon,h}^\star(t,s,\theta) = R_{1,N}(t,s,\theta) + R_{2,N}(t,s,\theta),$$

where

$$R_{1,N}(t,s,\theta) = \frac{1}{n}\sum_{i=1}^{n-h}\left(\frac{X_i^2(t)}{\tilde{\sigma}_i^2(t,\theta)} - 1\right)\left(\frac{X_{i+h}^2(s)}{\tilde{\sigma}_{i+h}^2(s,\theta)} - 1\right) - \left(\frac{X_i^2(t)}{\tilde{\sigma}_i^2(t,\theta)} - 1\right)\left(\frac{X_{i+h}^2(s)}{\sigma_{i+h}^2(s,\theta)} - 1\right)$$

and

$$R_{2,N}(t,s,\theta) = \frac{1}{n}\sum_{i=1}^{n-h}\left(\frac{X_i^2(t)}{\tilde{\sigma}_i^2(t,\theta)} - 1\right)\left(\frac{X_{i+h}^2(s)}{\sigma_{i+h}^2(s,\theta)} - 1\right) - \left(\frac{X_i^2(t)}{\sigma_i^2(t,\theta)} - 1\right)\left(\frac{X_{i+h}^2(s)}{\sigma_{i+h}^2(s,\theta)} - 1\right).$$

We note that Assumption B.6 implies that $\tilde{\sigma}_i^2(t,\theta) > \delta$ and $\sigma_i^2(t,\theta) > \delta$ uniformly in t and $\theta \in \Theta$, hence with this, the triangle inequality, and simple arithmetic yields that

$$\begin{aligned}
|R_{1,N}(t,s,\theta)| &\leq \frac{1}{n}\sum_{i=1}^{n-h}\left|\frac{X_i^2(t)}{\tilde{\sigma}_i^2(t,\theta)} - 1\right| |X_{i+h}^2(s)| \left|\frac{\tilde{\sigma}_{i+h}^2(s,\theta) - \sigma_{i+h}^2(s,\theta)}{\tilde{\sigma}_{i+h}^2(s,\theta)\sigma_{i+h}^2(s,\theta)}\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n-h}\left|\left(\frac{X_i^2(t)}{\tilde{\sigma}_i^2(t,\theta)} - 1\right) X_{i+h}^2(s)\right| \left|\frac{\tilde{\sigma}_{i+h}^2(s,\theta) - \sigma_{i+h}^2(s,\theta)}{\tilde{\sigma}_{i+h}^2(s,\theta)\sigma_{i+h}^2(s,\theta)}\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n-h}\left|\frac{X_i^2(t)X_{i+h}^2(s)}{\delta} - X_{i+h}^2(s)\right| \left|\frac{\tilde{\sigma}_{i+h}^2(s,\theta) - \sigma_{i+h}^2(s,\theta)}{\delta^2}\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n-h}\left|\frac{X_i^2(t)X_{i+h}^2(s)}{\delta} - X_{i+h}^2(s)\right| \frac{\|\tilde{\sigma}_{i+h}^2(\cdot,\theta) - \sigma_{i+h}^2(\cdot,\theta)\|_\infty}{\delta^2} \\
&\leq \frac{1}{n}\sum_{i=1}^{n-h}\left|\frac{X_i^2(t)X_{i+h}^2(s)}{\delta} - X_{i+h}^2(s)\right| \frac{c_1\rho^{i+h}}{\delta^2}, \quad a.s..
\end{aligned}$$

It follows from Assumption B.1 and the proof of Theorem 2.1 in Aue et al. (2017) that $E\|\sigma_i^2(\cdot,\theta_0)\|^2 < \infty$. Now using the Cauchy-Schwarz inequality, the stationarity of the solution X_i, the fact that σ_i^2 is measurable with respect to $\mathcal{F}_{i-1}$, and Assumption B.5, we have that

$$E\|X_i^2(\cdot)X_{i+h}^2(\cdot)\| = E\|X_i^2\|\|X_{i+h}^2\| \le E\|X_i^2\|^2 = E\int \sigma_i^4(t)\varepsilon_i^4(t)dt \le E\|\varepsilon_0^4\|_\infty E\|\sigma_i^2\|^2 < \infty.$$

From this it follows that $E\|X_i^2(\cdot)X_{i+h}^2(\cdot)/\delta - X_{i+h}^2(\cdot)\| < c_2$, for a positive constant c_2, and hence

$$\sup_{\theta\in\Theta}\sqrt{N}E\|R_{1,N}(\cdot,\cdot,\theta)\| \le \frac{1}{\sqrt{N}}\sum_{i=1}^{n-h}\frac{c_1c_2}{\delta^3}\rho^{i+h} = o(1).$$

We therefore have by Markov's inequality that $\sup_{\theta\in\Theta}\sqrt{N}\|R_{1,N}(\cdot,\cdot,\theta)\| = o_P(1)$. It follows similarly that $\sup_{\theta\in\Theta}\sqrt{N}\|R_{2,N}(\cdot,\cdot,\theta)\| = o_P(1)$, which establishes (B.7). In order to show (B.8), we first note that by simply differentiating (B.2) that with $\mathbb{1}^{(j)}$ denoting an L vector of zeros with a single 1 in the j'th position, and $\mathbb{1}^{(j,k)}$ being a $L\times L$ matrix of zeroes with a single 1 in the $(j,\ell)'th$ position, that for $1\le j,k\le L$,

$$\frac{\partial\sigma_i^2(t,\theta)}{d_j} = \left(\sum_{\ell=0}^{\infty}B^\ell\mathbb{1}^{(j)}\right)^\top\mathbf{\Phi}(t),\quad \frac{\partial\sigma_i^2(t,\theta)}{a_{j,k}} = \left(\sum_{\ell=0}^{\infty}B^\ell\mathbb{1}^{(j,k)}\mathbf{x}_{i-1-\ell}\right)^\top\mathbf{\Phi}(t),$$

and

$$\frac{\partial\sigma_i^2(t,\theta)}{b_{j,k}} = \left(\sum_{\ell=0}^{\infty}\left\{\sum_{r=1}^{\ell}B^{r-1}\mathbb{1}^{(j,k)}B^{\ell-r}\right\}\xi_i\right)^\top\mathbf{\Phi}(t).$$

Similarly

$$\text{(B.10)}\qquad \frac{\partial\tilde\sigma_i^2(t,\theta)}{d_j} = \left(\sum_{\ell=0}^{i-1}B^\ell\mathbb{1}^{(j)}\right)^\top\mathbf{\Phi}(t),\quad \frac{\partial\tilde\sigma_i^2(t,\theta)}{a_{j,k}} = \left(\sum_{\ell=0}^{i-1}B^\ell\mathbb{1}^{(j,k)}\mathbf{x}_{i-1-\ell}\right)^\top\mathbf{\Phi}(t),$$

and

$$\text{(B.11)}\qquad \frac{\partial\tilde\sigma_i^2(t,\theta)}{b_{j,k}} = \left(\sum_{\ell=0}^{i-1}\left\{\sum_{r=1}^{\ell}B^{r-1}\mathbb{1}^{(j,k)}B^{\ell-r}\right\}\xi_i\right)^\top\mathbf{\Phi}(t).$$

By Assumption B.4 it follows similarly as (B.9) that

$$\max_{j\in\{1,\ldots,L+2L^2\}}\sup_{\theta\in\Theta}\left\|\frac{\partial\tilde{\sigma}_i^2(\cdot,\theta)}{\partial\theta^{(j)}}-\frac{\partial{\sigma^2}_i(\cdot,\theta)}{\partial\theta^{(j)}}\right\|_\infty\leq c_4\rho^i,$$

for a $0<\rho<1$. From this (B.8) follows similarly as (B.7), and so we omit the details. $\square$

Proof of Theorem 3.1. The proof of Theorem 3.1, is inspired by the proof of Theorem 8.2 in Francq and Zakoïan, (2010). Noting that $\hat{\gamma}_{\varepsilon,h}(t,s)=\tilde{\gamma}_{\varepsilon,h}(t,s,\hat{\theta}_N)$, we get by applying a one term Taylor's expansion centered at θ_0 that for all $t,s\in[0,1]$ and $1\leq h\leq K$,

$$\sqrt{N}\hat{\gamma}_{\varepsilon,h}(t,s)=\sqrt{N}\tilde{\gamma}_{\varepsilon,h}(t,s,\theta_0)+\frac{\partial\tilde{\gamma}_{\varepsilon,h}(t,s,\theta^*)}{\partial\theta}\sqrt{N}(\hat{\theta}_N-\theta_0), \tag{B.12}$$

where θ^* is $L+2L^2$ dimensional rectangle between θ_0 and $\hat{\theta}_N$. By Lemma B.1, there exists a function $R_{3,N}(t,s)$ satisfying that $\|R_{3,N}(\cdot,\cdot)\|=o_P(1)$, and

$$\begin{aligned}\sqrt{N}\tilde{\gamma}_{\varepsilon,h}(t,s,\theta_0)+\frac{\partial\tilde{\gamma}_{\varepsilon,h}(t,s,\theta^*)}{\partial\theta}\sqrt{N}(\hat{\theta}_N-\theta_0)=\sqrt{N}\gamma^\star_{\varepsilon,h}(t,s,\theta_0)+&\frac{\partial\gamma^\star_{\varepsilon,h}(t,s,\theta^*)}{\partial\theta}\sqrt{N}(\hat{\theta}_N-\theta_0)\\&+R_{3,N}(t,s).\end{aligned}$$

Let

$$\hat{G}_{N,h}(t,s,\theta)=\frac{\partial\gamma^\star_{\varepsilon,h}(t,s,\theta)}{\partial\theta},$$

so for each fixed θ, $\hat{G}_{n,h}:[0,1]^2\to\mathbb{R}^{L+2L^2}$. Calculating the derivative for each $t,s\in[0,1]$ yields that

$$\begin{aligned}\hat{G}_{N,h}(t,s,\theta)=&-\frac{1}{N}\sum_{i=1}^{N-h}\left(\frac{X_i^2(t)}{\sigma_i^4(t,\theta)}\frac{\partial\sigma_i^2(t,\theta)}{\partial\theta}\right)\left(\frac{X_{i+h}^2(s)}{\sigma_{i+h}^2(s,\theta)}-1\right)\\&-\frac{1}{N}\sum_{i=1}^{N-h}\left(\frac{X_i^2(t)}{\sigma_i^2(t,\theta)}-1\right)\left(\frac{X_{i+h}^2(s)}{\sigma_i^4(s,\theta)}\frac{\partial\sigma_{i+h}^2(s,\theta)}{\partial\theta}\right).\end{aligned} \tag{B.13}$$

Applying another Taylor's expansion to $\hat{G}_{N,h}$ centered at θ_0 gives that

$$\hat{G}_{N,h}(t,s,\theta^*)=\hat{G}_{N,h}(t,s,\theta_0)+\frac{\partial\hat{G}_{N,h}(t,s,\theta^{**})}{\partial\theta}(\theta^*-\theta_0),$$

where θ^{**} is between θ^* and θ_0. It follows as in the proof of Lemma B.1 that

$$\max_{j\in\{1,\ldots,L+2L^2\}} \sup_{\theta\in\Theta} E\left\|\frac{\partial \hat{G}_{N,h}(\cdot,\cdot,\theta)}{\partial\theta^{(j)}}\right\| < \infty,$$

and hence using the strong consistency of $\hat{\theta}_N$ we obtain that

$$\|\hat{G}_{N,h}(t,s,\theta^*) - \hat{G}_{N,h}(t,s,\theta_0)\| = o_P(1).$$

From (B.13), we see that

$$\begin{aligned}\hat{G}_{N,h}(t,s,\theta_0) = &-\frac{1}{N}\sum_{i=1}^{N-h}\left(\frac{\varepsilon_i^2(t)}{\sigma_i^2(t,\theta)}\frac{\partial\sigma_i^2(t,\theta)}{\partial\theta}\right)\left(\varepsilon_{i+h}^2(s)-1\right)\\ &-\frac{1}{N}\sum_{i=1}^{N-h}\left(\varepsilon_i^2(t)-1\right)\left(\frac{\varepsilon_{i+h}^2(s)}{\sigma_i^2(s,\theta)}\frac{\partial\sigma_{i+h}^2(s,\theta)}{\partial\theta}\right).\end{aligned}$$

Since σ_i^2 is $\mathcal{F}_{i-1}$ measurable and $E\varepsilon_{i+h}^2(s) = 1$, the expectation of the first term is zero so that,

$$E\hat{G}_{N,h}(t,s,\theta_0) = \frac{N-h}{N}G_h(t,s),$$

Further, since σ_i^2 is ergodic, we have by the ergodic theorem in Hilbert space see Bosq (2000) that

$$\max_{j\in\{1,\ldots,L+2L^2\}} \|\hat{G}_{N,h}^{(j)}(\cdot,\cdot,\theta_0) - G_h^{(j)}\| = o_P(1).$$

Combining these results with (B.12), we see that

$$\sqrt{N}\hat{\gamma}_{\varepsilon,h}(t,s) = \sqrt{N}\gamma^{\star}_{\varepsilon,h}(t,s,\theta_0) + G_h(t,s)\sqrt{N}(\hat{\theta}_N - \theta_0) + R_{4,N}(t,s), \tag{B.14}$$

where $\|R_{4,N}\| = o_P(1)$. We note that

$$\sqrt{N}\gamma^{\star}_{\varepsilon,h}(t,s,\theta_0) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N-h}(\varepsilon_i^2(t)-1)(\varepsilon_{i+h}^2(s)-1),$$

depends solely on the error process: in particular it is $\sqrt{N}$ times the estimated autocovariance of the squared error processes that was considered in Appendix A. Let
$\Gamma^{(\varepsilon)}_{N,K} = (\sqrt{N}\gamma^{\star}_{\varepsilon,h}(\cdot,\cdot,\theta_0), ..., \sqrt{N}\gamma^{\star}_{\varepsilon,K}(\cdot,\cdot,\theta_0))^{\top}$ and
$\Gamma^{(\theta)}_{N,K} = (G_1(\cdot,\cdot)\sqrt{N}(\hat{\theta}_N - \theta_0), ..., G_h(\cdot,\cdot)\sqrt{N}(\hat{\theta}_N - \theta_0))^{\top}$. It follows then from (B.14) that

$$\|\Gamma^{(\varepsilon,\theta)}_{N,K} - (\Gamma^{(\varepsilon)}_{N,K} + \Gamma^{(\theta)}_{N,K})\|_{\mathcal{G},1} = o_P(1).$$

We now aim at establishing the weak limit of $\Gamma^{(\varepsilon)}_{N,K} + \Gamma^{(\theta)}_{N,K}$ in $\mathcal{G}_1$. $\Gamma^{(\varepsilon)}_{N,K}$ is tight in $\mathcal{G}_1$ as was established in Appendix A, and $\Gamma^{(\theta)}_{N,K}$ is tight in $\mathcal{G}_1$ since $\mathbb{R}^{L+2L^2}$ is sigma-compact, hence $\Gamma^{(\varepsilon)}_{N,K} + \Gamma^{(\theta)}_{N,K}$ is tight in $\mathcal{G}_1$. According to the proof of Theorem 4 on pg 19 of Aue et al. (2017), in particular their equations 5.15–5.22, we have under Assumptions B.1–B.6 that

$$\left\| \sqrt{N}(\hat{\theta}_N - \theta_0) - \frac{Q_0^{-1}}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial \mathfrak{s}_i^2(\theta_0)^{\top}}{\partial \theta} [\mathbf{x}_i^2 - \mathfrak{s}_i^2(\theta_0)] \right\|_E = o_P(1).$$

Therefore if $z \in \mathcal{G}_1$,

$$\begin{aligned} \text{(B.15)} \qquad \langle \Gamma^{(\varepsilon)}_{N,K} + \Gamma^{(\theta)}_{N,K}, z \rangle_{\mathcal{G},1} &= \frac{1}{\sqrt{N}} \left\{ \sum_{i=1}^{N} \left[\sum_{h=1}^{K} \left(\langle (\varepsilon_i^2(\cdot) - 1) \otimes (\varepsilon_{i+h}^2(\cdot) - 1), z^{(h)} \rangle \right.\right.\right. \\ &\qquad \left.\left.\left. + \langle G_h, z^{(h)} \rangle_* Q_0^{-1} \frac{\partial \mathfrak{s}_i^2(\theta_0)^{\top}}{\partial \theta} [\mathbf{x}_i^2 - \mathfrak{s}_i^2(\theta_0)] \right) \right] \right\} \\ &=: \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \nu_i(z), \end{aligned}$$

where $\langle G_h, z^{(h)} \rangle_*$ is used to denote that the inner-product is carried out coordinate-wise, so that $\langle G_h, z^{(h)} \rangle_* \in \mathbb{R}^{L+2L^2}$. Noting that 1) $\mathfrak{s}_i^2$ is $\mathcal{F}_{i-1}$ measurable, and 2) $E[\mathbf{x}_i - \mathfrak{s}_i(\theta_0)|\mathcal{F}_{i-1}] = 0$, we see that $\nu_i(z)$ form a martingale difference sequence. Moreover, $\nu_i(z)$, $i \in \mathbb{Z}$ is a stationary sequence since $(\varepsilon_i^2, \mathfrak{s}_i^2, \mathbf{x}_i^2)$ is stationary. Using the Cauchy Schwarz inequality, it follows readily that for a positive constant c_5

$$E\nu_i^2(z) \leq c_5 \left\{ \|z\|_{\mathcal{G},1} E\|\varepsilon_i^2\|^2 + \max_{1\leq h\leq K} E\left(\langle G_h, z \rangle_* Q_0^{-1} \frac{\partial \mathfrak{s}_i^2(\theta_0)^{\top}}{\partial \theta} [\mathbf{x}_i^2 - \mathfrak{s}_i^2(\theta_0)] \right)^2 \right\} < \infty,$$

using Assumption B.5. Hence by the martingale central limit theorem, see for example Corrollary A.1 in Francq and Zakoïan (2010),

$$\langle \Gamma^{(\varepsilon)}_{N,K} + \Gamma^{(\theta)}_{N,K}, z\rangle_{\mathcal{G},1} \xrightarrow{\mathcal{D}} N(0, \text{Var}(\nu_0(z))).$$

Straightforward calculation shows that

$$\text{Var}(\nu_0(z)) = \langle \Psi_{K,\varepsilon,\theta}(z), z\rangle_{\mathcal{G},1},$$

which establishes the first half of the Theorem. The asymptotic distribution of the $M_{N,K,\varepsilon}$ follows immediately from the continuous mapping theorem and the Karhunen-Loéve representation.

□

B.1. **Estimation of $\Psi^{(\varepsilon,\theta)}_{K,h,g}$.** Estimating the covariance kernel $\Psi^{(\varepsilon,\theta)}_{K,h,g}$ for $1 \le h \le K$, $1 \le g \le K$ is nontrivial given its dimensionality and its complicated composition. We describe here how to numerically estimate the key components in $\hat{\Psi}^{(\varepsilon,\theta)}_{K,h,g}$.

Recall from (3.8), we focus on the estimation of last three terms because $C_\varepsilon(t,s,u,v)$ can be estimated straightforwardly. In particular, it is not hard to see that to estimate $C^{\varepsilon,\theta}_{h,g}(t,s,u,v)$ and $G_h^\top(t,s)Q_0^{-1}H_0^\top J_0 H_0 Q_0^{-1} G_g(u,v)$, we need to estimate $G_h(t,s)$ and $\frac{\partial \mathfrak{s}_0^2(\theta_0)}{\theta}$. Note that $\hat{C}^{\varepsilon,\theta}_{g,h}(u,v,t,s)$ does not need to be estimated but is simply obtained by transposing of $\hat{C}^{\varepsilon,\theta}_{h,g}(t,s,u,v)$.

In order to estimate $G_h(t,s)$, we consider the partial derivative

$$\frac{\partial \hat{\sigma}^2_{i+h}(s,\hat{\theta}_N)}{\partial \theta} = \frac{\partial \sum_j^L \hat{\mathfrak{s}}^2_{i+h,j}(\hat{\theta}_N)\hat{\phi}_j(s)}{\partial \theta} = [\, \frac{\partial \hat{\mathfrak{s}}^2_{i+h}(\hat{\theta}_N)}{\partial d_1}\hat{\phi}_1(s), \ldots, \frac{\partial \hat{\mathfrak{s}}^2_{i+h}(\hat{\theta}_N)}{\partial d_L}\hat{\phi}_L(s), \frac{\partial \hat{\mathfrak{s}}^2_{i+h}(\hat{\theta}_N)}{\partial a_{1,1}}\hat{\phi}_1(s)\hat{\phi}_1(t)$$
$$\ldots, \frac{\partial \hat{\mathfrak{s}}^2_{i+h}(\hat{\theta}_N)}{\partial a_{L,L}}\hat{\phi}_L(s)\hat{\phi}_L(t), \frac{\partial \hat{\mathfrak{s}}^2_{i+h}(\hat{\theta}_N)}{\partial b_{1,1}}\hat{\phi}_1(s)\hat{\phi}_1(t), \ldots, \frac{\partial \hat{\mathfrak{s}}^2_{i+h}(\hat{\theta}_N)}{\partial b_{L,L}}\hat{\phi}_L(s)\hat{\phi}_L(t)\,]^\top.$$

where $\hat{\phi}_1, \ldots, \hat{\phi}_L$ are the estimated principal components. Considering that Equation (3.2) can be written as $\mathfrak{s}_i^2 = \sum_{l=0}^\infty B^l(D + A\mathbf{x}^2_{i-l-1})$, we use (B.10) and (B.11) to obtain,

$$\frac{\partial \hat{\mathfrak{s}}_i^2(\hat{\theta}_N)}{\partial d_j} = \sum_{\ell=0}^\infty \hat{B}^\ell \mathbb{1}^{(j)}, \quad \frac{\partial \hat{\mathfrak{s}}_i^2(\hat{\theta}_N)}{\partial a_{j,k}} = \sum_{\ell=0}^\infty \hat{B}^\ell \mathbb{1}^{(j,k)}\hat{\mathbf{x}}^2_{i-\ell-1},$$
$$\frac{\partial \hat{\mathfrak{s}}_i^2(\hat{\theta}_N)}{\partial b_{j,k}} = \sum_{\ell=1}^\infty [\sum_{i=1}^\ell \hat{B}^{i-1}\ell^{(j,k)}\hat{B}^{\ell-i}](\hat{D} + \hat{A}\hat{\mathbf{x}}^2_{i-\ell-1}), \quad 1 \le j \le L, \quad 1 \le k \le L. \tag{B.16}$$

We choose the initial value of $\mathbf{x}_0 = \hat{D}$. As $\hat{B}^\ell$ decays geometrically at a fast rate under the condition of stationarity, we only consider $0 \le \ell \le 5$ in the implementation for computational efficiency, and further lags beyond 5 are ignored. Since $\frac{\partial \hat{\sigma}_i^2(s,\hat{\theta}_N)}{\partial \theta}$ is a $(L + 2L^2) \times 1$ vector, we consider the column sums on each individual partial derivative.

Similarly to Equation (B.16), we are able to estimate $\frac{\partial \mathfrak{s}_0^2(\theta_0)}{\theta}$, which is a critical component in estimating H_0 and Q_0. Note that $\frac{\partial \mathfrak{s}_0^2(\theta_0)}{\theta}$ is a $(L + 2L^2) \times L$ matrix, so that we do not need to consider the column sums. Thus, H_0 and Q_0 can be estimated by taking their sample means.

Lastly, we show how to adapt the above estimations into a FARCH(1) model in the simulation section. The number of parameters in the FARCH(1) model is $L + L^2$ because the GARCH coefficient matrix B vanishes in this case. Thus, we have $\mathfrak{s}_i^2 = D + A\mathbf{x}_{i-1}^2$, and the derivatives in (B.16) only act on coefficients D and A and depends on previous values at lag 1.

References

Aue, A., L. Horváth, L., Pellatt, D. (2017). Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis 38*, 3–21.

Barndorff-Nielsen, O. E., Shephard, N. (2004). Econometric analysis of realized co-variance: High frequency based covariance, regression and correlation in financial economics. *Econometrica 72*, 885–925.

Bauwens, L., Laurent, S., Rombouts, J. V. K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics 21*, 79–109.

Bogachev V. (1998). *Gaussian Measures*. American Mathematical Society.

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

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer.

Boussama, F., Fuchs, F., Stelzer. R. (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and Their Applications 121*, 2331–2360.

Brockwell, P. J., Richard, A. D. (1991). *Time Series: Theory and Methods*. Springer.

Catani, P., Teräsvirta, T., Yin, M. (2017). A Lagrange multiplier test for testing the adequacy of constant conditional correlation GARCH model. *Econometric Reviews*, *36*, 599–621.

Cerovecki, C., Francq, C., Hörmann, S., Zakoïan, J. (2019). Functional GARCH models: the quasi-likelihood approach and its applications. *Journal of Econometrics*, *209*, 353–375.

Chiou, J-M., Müller, H-G. (2007). Diagnostics for functional regression via residual processes. *Computational Statistics & Data Analysis 51*, 4849–4863.

Duchesne P., Lalancette, S. (2003). On testing for multivariate ARCH effects in vector time series models. *Canadian Journal of Statistics 31*, 275–292.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society 50*, 987–1007.

Engle, R. F., Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*, 122–150.

Francq, C., Zakoïan, J. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Gabrys, R., Horváth, L., Kokoszka, P. (2010). Tests for error correlation in the functional linear model. *Journal of the American Statistical Association 105*, 1113–1125.

Gabrys, R., Kokoszka, P. (2007). Portmanteau test of independence for functional observations. *Journal of the American Statistical Association 102*, 1338–1348.

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. *The Annals of Statistics 38*, 1845–1884.

Horváth, L., Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer.

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

Horváth, L., Hušková, M., Rice, G. (2013). Test of independence for functional data. *Journal of Multivariate Analysis 117*, 100–119.

Kokoszka, P., Reimherr, M. (2013). Predictability of shapes of intraday price curves. *The Econometrics Journal 16*, 285–308.

Kokoszka, P., Rice, R., Shang, H. L. (2017). Inference for the autocovariance of a functional time series under conditional heteroscedasticity. *Journal of Multivariate Analysis 162*, 32–50.

Koopman, S. J., Jungbacker, B., Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance 12*, 445–475.

Kroner, K. F., Ng, V. (1998). Modeling asymmetric comovements of asset returns. *The Review of Financial Studies 11*, 817–844.

Li, W. K. (2003). *Diagnostic Checks in Time Series*. Chapman and Hall/CRC.

Li, W. K., Mak, T. K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis 15*, 627–636.

Ljung, G., Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometricka 66*, 67–72.

McLeod, A. I., Li, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis 4*, 269–273.

Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*. SIAM.

Müller, H-G., Sen, R., Stadtmüller, U. (2011). Functional data analysis for volatility. *Journal of Econometrics 165*, 233–245.

Ramsay, J. O., Silverman, B. W. (2006). *Functional Data Analysis*. Wiley Online Library.

Riesz, F., Sz.-Nagy, B. (1990). *Functional Analysis*. Dover.

Tse Y. K., Tsui, A. (1999). A note on diagnosing multivariate conditional heteroscedasticity models. *Journal of Time Series Analysis 20*, 679–691.

Tse, Y. K. (2002). Residual-based diagnostics for conditional heteroscedasticity models. *The Econometrics Journal 5*, 358–373.

Zhang, J. T. (2013). *Analysis of Variance for Functional Data*. Chapman & Hall. London.

Zhang, X. (2016). White noise testing and model diagnostic checking for functional time series. *Journal of Econometrics 194*, 76–95.