Exploring volatility of crude oil intra-day return curves: a functional GARCH-X model

Rice, Gregory and Wirjanto, Tony and Zhao, Yuqian

University of Waterloo, University of Waterloo, University of Essex

18 August 2021

EXPLORING VOLATILITY OF CRUDE OIL INTRA-DAY RETURN CURVES: A FUNCTIONAL GARCH-X MODEL

GREGORY RICE[§], TONY WIRJANTO[‡] AND YUQIAN ZHAO[†]

Abstract. Crude oil intra-day return curves collected from the commodity futures market often appear to be serially uncorrelated and long-range dependent. Existing functional GARCH models, while able to accommodate short range conditional heteroscedasticity, are not designed to capture long-range dependence. We propose and study a new functional GARCH-X model for this purpose, where the covariate X is chosen to be weakly stationary and long-range dependent. Functional analogs of autocorrelation coefficients of squared processes for this model are derived, and compared to those estimated from crude oil return curves. The results show that the FGARCH-X model provides a significant correction to existing functional volatility models in terms of an in-sample fitting, while its out-of-sample performances do not appear to be more superior than the existing functional GARCH(1,1) model.

JEL Classification: C13, C32, C58, G10, G17

Keywords: Crude oil intra-day return curves, volatility modeling and forecasting, functional GARCH-X model, long-range dependence, basis selection.

* We thank participants at the SIAM Conference on Financial Mathematics & Engineering, the 2020 NBER-NSF Bayesian Inference in Econometrics and Statistics (SBIES) conference, the CMStatistics 2020 conference, and the ECMFE seminar at the university of Essex for helpful comments.

[§]Department of Statistics and Actuarial Science, University of Waterloo, Canada. Email: grice@uwaterloo.ca.
[‡]Department of Statistics and Actuarial Science, University of Waterloo, Canada. Email: twirjanto@uwaterloo.ca.
[†]**Corresponding Author**. Essex Business School, University of Essex, UK. Email: y.zhao@essex.ac.uk.

1. Introduction

Modeling and forecasting the volatility of crude oil returns are of great interest to market practitioners, researchers and policy makers, given that it is often used as a leading indicator for risk, portfolio management and option pricing. Previous empirical works have typically applied scalar GARCH-type processes (Engle, 1982; Bollerslev, 1986) to model and predict the crude oil volatility, see for instance Wei et al. (2010) and Charles and Darné (2014). More recently, given the increasing availability of high (in time) resolution data on financial assets, one is often faced with the question of how to incorporate such data in order to improve the crude oil volatility modeling and forecasting. For empirical attempts to model crude oil volatility with intra-day data, we refer readers to Ma et al. (2019) and Zhang and Wang (2019).

Methodologically, early work (Andersen and Bollerslev, 1997) has found that intra-day returns behave quite distinctly from inter-daily returns. This renders scalar GARCH-type models less effective when they are used in a high frequency environment. As a consequence, several studies have subsequently focused on exploring the usefulness of information stored at an intra-day level for inter-daily volatility modeling in financial econometrics (e.g., Bollerslev et al., 2016; Gorgi et al., 2019). In the mean time, intra-day trading activities call for risk management tools that can be suitably used at the intra-day level, which encourages modeling volatility dynamics directly at the intra-day level. To this end, in the current paper we consider the modeling of the volatility of crude oil intra-day return curves.

Analyzing intra-day return curves has recently been carried out by using techniques developed in functional data analysis. Daily return curves may be constructed by interpolating or smoothing high resolution, even tick-by-tick, data, which has the benefit of preserving all of the intra-day price information. The resulting time series of daily curves can then be analyzed by using techniques developed in functional time series analysis. We refer the readers to Bosq (2000) and Horváth and Kokoszka (2012) for monographs on functional data analysis and its extension to functional time series; and Kearney and Shang (2019) and Rice et al. (2020b) for financial applications involving these return curves. There are very few studies that explore the volatility of crude oil intra-day return curves, providing us with an opportunity to bridge between the methodological development and its potential applications in the financial markets.

In this paper, we model and forecast conditional volatility of West Texas Intermediate (WTI) crude oil intra-day return curves collected from the commodity futures market. By investigating this curve process, we found that crude oil return curves appear to be serially uncorrelated, but also exhibit long-range conditional heteroscedastic. Given the fact that the existing functional conditional volatility models (i.e., functional ARCH model – Hörmann et al., 2013; functional GARCH(1,1) model – Aue et al., 2017; functional GARCH(p,q) model – Cerovecki et al., 2019) are all designed for modeling short-range conditional heteroscedasticity, we propose a parsimonious functional GARCH-X (FGARCH-X) model, where the stationary covariate X is chosen to accommodate long-range conditional heteroscedasticity in the intra-day return curves. The weak stationary solution and the autocorrelation structure of the new model process are provided in this paper. In addition, we introduce two new basis selection methods in the dimension reduction for the purpose of estimating the models, namely, a functional sparse and non-negative basis and a truncated predictive factors. A Monte Carlo simulation study shows that compared with the functional ARCH (FARCH) and functional GARCH (FGARCH) processes, the proposed FGARCH-X model can better restore the autocorrelation dynamics observed from the real data.

In the application part, we use the FGARCH-type models – FGARCH(1,1) and FARCH(5), as well as the proposed FGARCH-X model to fit and predict conditional volatility of crude oil return curves. We consider four types of exogenous covariates derived from historical intra-day returns and volumes for the FGARCH-X model. In order to evaluate the out-of-sample forecast, we compute intra-day Value-at-Risk (VaR) through the predicted values of conditional volatility. Our results indicate that in terms of the in-sample fitting, all three model candidates pass a series of goodness-of-fit tests and are able to accommodate conditional heteroscedasticity of crude oil intra-day return curves, but the FGARCH-X model better restores the autocorrelation structure in the squared process of return curves. In terms of the out-of-sample performance, all of the models are able to forecast valid intra-day VaR by allowing their violation processes to pass the unbiased and independence backtests (Rice et al. 2020b). Despite its ability of explaining long-range conditional heteroscedasticity, the FGARCH-X model does not appear to offer a clear-cut superiority over the FGARCH(1,1) model. This result extends the findings of Hansen and Lunde (2005) to functional conditional volatility models.

Our paper contributes to the existing literature in three important aspects. First, we introduce a novel way for modeling the volatility of crude oil intra-day returns, which refines the intra-day risk management tool in commodity futures markets. Second, there have been only a few studies that focus on the importance of modeling the second-moment dynamics of functional time series data (Hörmann et al., 2013; Aue et al., 2017; Cerovecki et al., 2019). We observe that the intra-day return curves derived from crude oil market are second-order long-range dependent and develop a simple model to enhance the ability of the model to capture this stylized fact. Finally, we address the inherent problem in the estimation of functional volatility models by proposing data-driven, non-negative and predictive basis functions for a dimension reduction.

The remaining parts of this paper are organized as follows. Section 2 studies the characteristics of the crude oil intra-day price data. Section 3 proposes a functional GARCH-X model and discusses the selection of basis functions for the functional volatility model estimation. Section 4 investigates the autocorrelation structure of the proposed functional volatility models by means of a Monte Carlo simulation. Lastly we carry out a forecasting exercise for conditional volatility of crude oil futures intra-day returns in Section 5. Section 6 concludes the paper with a few remarks.

2. Main features of crude oil intra-day return curves

This section introduces the data and studies the stylish features of intra-day return curves. We collect the raw price data at a five-minute frequency on WTI crude oil contracts quoted on the NYMEX-CME. The price time series is obtained by obeying the rule that uses the front-month future contract, and rolls over when the next contract becomes more traded than the expiring one, normally few days before the expiration. Investors are able to trade crude oil around the clock from Sunday to Friday. To preserve the quality of the data we only use the main market time zone from 9:00 to 14:30, which delivers 66 grid points at the intra-day level. Our sample covers the period from 2nd January 2015 to 1st May 2020, including 1,375 trading days.

In order to construct intra-day return curves, we denote $P_i(t_j)$ as the price of crude oil on day i at intra-day time t_j. After applying interpolation or smoothing techniques to these raw data, see e.g. Chapter 3 of Ramsay and Silverman (2006), full intra-day price curves $P_i(t)$ can be produced, where we assume that the intra-day time parameter t is normalized to the unit interval $t \in [0, 1]$. We then define the weakly stationary overnight cumulative intra-day log

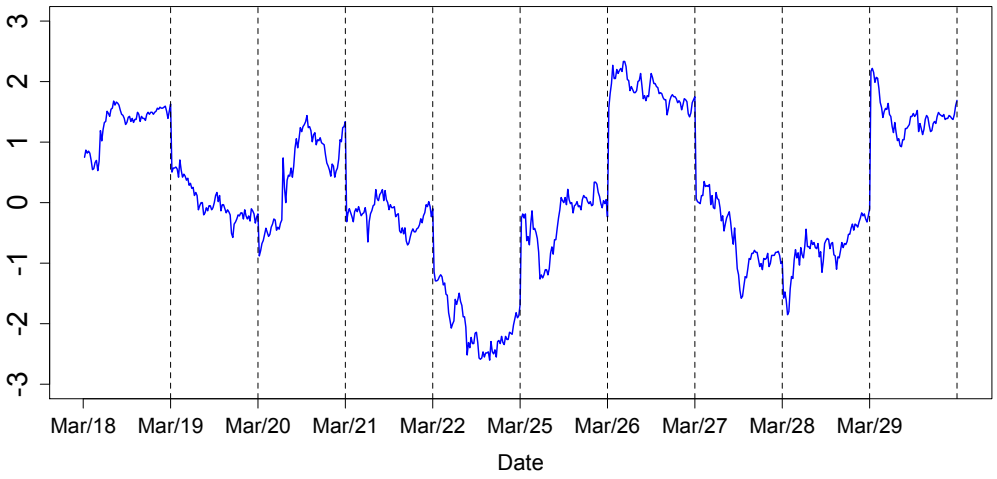
return (OCIDR) curves as

$$Y_i(t) = 100[\log P_i(t) - \log P_{i-1}(1)], \quad 1 \leq i \leq N, \quad t \in [0, 1] \tag{2.1}$$

where $P_{i-1}(1)$ is the adjusted closing price from the previous trading day. The OCIDR curve was initially introduced in Rice et al. (2020a). Compared with other versions of intra-day return curves (Kokoszka and Reimherr, 2013; Kokoszka et al., 2017), the OCIDR curves are more suitable for modeling volatility, given the important role of the overnight effect, which they capture (Hansen and Lunde, 2006).

Figure 2.1 displays an example plot of the OCIDR curves of crude oil from March 2019 constructed from 5 minute resolution price data. The basic properties of OCIDR curves can

FIGURE 2.1. Plots of ten OCIDR cruves derived from five-minute resolution records of WTI crude oil from March, 2019.



be investigated by implementing recently developed hypothesis tests for functional time series data, including the KPSS-typed stationary test (Horváth et al. 2014), the normality test based on static FPCA (Górecki et al., 2018), the independent test (Kokoszka et al. 2017), and the heteroscedasticity test (Rice et al., 2020a). The technical details of each of these tests are omitted from presentation and readers are referred to the respective papers. Overall, we find that the crude oil OCIDR curves are weakly stationarity, non-normal, serially uncorrelated but conditionally heteroscedastic. For detailed p-values of these tests, see Table 5.1 in Section 5.

In modeling curve process with such features, Hörmann et al. (2013) adapted the idea of the ARCH model to the functional time series setting, in which conditional volatility of intra-day

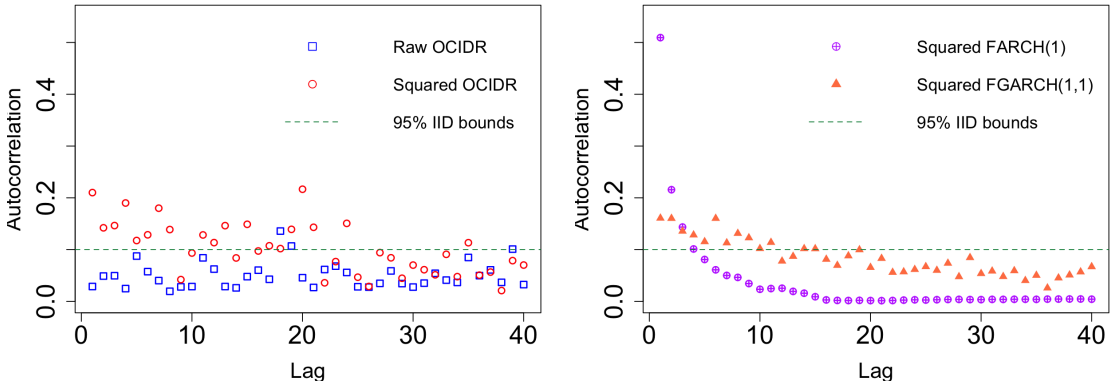
return curves is modeled linearly in terms of the past squared curve. Their model was then subsequently generalized to FGARCH models in Aue et al. (2017) and Cerovecki et al. (2019). A well-known characteristic of daily asset returns data is the presence of long memory in the volatility process (Ding et al., 1993), and such a feature might be expected to appear in intra-day return curves as well.

In order to measure the autocorrelation structure of the observed return curves, we let $C_h(t,s) = \text{Cov}(Y_i(t), Y_{i+h}(s))$ and define a scale free functional autocorrelation measure (FACF) of the OCIDR curves at a non-negative lag h as

$$\rho_h = \frac{\left(\int_0^1 C_h^2(t,s)dtds\right)^{1/2}}{\int_0^1 C_0(t,t)dt} = \frac{\|C_h\|}{trace(C_0)}. \tag{2.2}$$

It is also of interest to consider the same quantity with Y_i being replaced by Y_i^2, in order to measure the extent of serial dependence in the squared OCIDR curves. Estimators of ρ_h as a function of h up to lag 40 for the raw and squared OCIDR curves derived from 5-minute resolution WTI crude oil prices are illustrated in the left panel of Figure 2.2, along with 95% confidence intervals for ρ_h centered at 0 constructed under the assumption that the curves are independent and identically distributed. Values of the estimators that are not included in these confidence intervals can be taken as evidence of significant autocorrelation in the sequence of raw or squared OCIDR curves. One can readily observe here that serial correlation estimated in the raw OCIDR curves is generally consistent with the series evolving as a (weak) white noise, but that the autocorrelation observed in the squared OCIDR curves decays quite slowly, indicating the presence of long memory in the volatility process. A similar finding is also evidenced by Casas and Gao (2008), who showed long-range dependence in the daily volatility functional time series. The right hand panel of Figure 2.2 shows the theoretical FACF of squared FARCH(1) and FGARCH(1,1) processes fit to the OCIDR curves, where it is apparent that the FGARCH(1,1) model appears to capture the short range dependence well, but does not appear to capture the same rate of decay of the FACF at long lags. To the best of our knowledge, no model has been developed to date to capture this stylized feature.

FIGURE 2.2. Left hand panel: Plots of estimators of ρ_h of the raw and squared OCIDR curves derived from 5-minute resolution WTI crude oil prices, along with a 95% confidence interval for these estimators constructed under the assumption that the series follows a strong white noise. Right hand panel: Theoretical approximations of the FACF of the squared FARCH and FGARCH fit using pseudo-maximum likelihood estimation to the OCIDR curves, with with the same 95% confidence interval as in the left hand figure for reference.



3. FUNCTIONAL VOLATILITY MODELS AND THE BASIS FUNCTION SELECTION

This section proposes an FGARCH-X model to account for the long-range second moment dependence of the crude oil return curves, and we also discuss the basis selection in the dimension reduction for the purpose of model estimation. For notational convenience, the intra-day return curve $y_i(t)$ is assumed to be square integrable and drawn from a $L^2[0,1]$ Hilbert space, equipped with an inner product $\langle y_1, y_2\rangle = \int y_1(t)y_2(t)dt$, resulting a norm $||y(t)|| = [\int y^2(t)dt]^{1/2}$, where $\int \equiv \int_0^1$. The kernel integral operator $\mathbf{g}(y)(t) = \int g(t,s)y(s)ds$, for $g(t,s) \in L^2[0,1]^2$. Considering that functional volatility processes are strictly positive, we further define subspaces $L^2[0,1]^+ = \{y \in L^2[0,1], y(t) \geq 0\}$ and $L^2[0,1]^+_* = \{y \in L^2[0,1], y(t) > 0\}$, for almost every $t \in [0,1]$.

3.1. **Functional GARCH-X model.**

Let us first focus on the FGARCH(p,q) model (Cerovecki et al., 2019), which accommodates conditional heteroscedasticity of the OCIDR curves by assuming that $y_i(t)$ follows a recursion relationship as specified below:

$$\begin{aligned} y_i(t) &= \sigma_i(t)\varepsilon_i(t), \quad t \in [0,1] \\ \sigma_i^2(t) &= \omega(t) + \sum_{j=1}^{q} \boldsymbol{\alpha}_j(y_{i-j}^2)(t) + \sum_{k=1}^{p} \boldsymbol{\beta}_k(\sigma_{i-k}^2)(t), \end{aligned} \tag{3.1}$$

where $\sigma_i^2(t)$ is a latent conditional volatility function, $\omega(t) \in L^2[0,1]_*^+$, and the kernels of the coefficient operators, $\alpha_j(t,s)$ and $\beta_k(t,s)$ are elements in $L^2[0,1]^+ \times L^2[0,1]^+$.

In addition, we also consider an FARCH(q) model, in which the conditional volatility equation is specified as,

$$\sigma_i^2(t) = \omega(t) + \sum_{j=1}^{\mathfrak{q}} \boldsymbol{\alpha}_j(y_{i-j}^2)(t). \tag{3.2}$$

Intuitively, increasing the lag length of the ARCH component in the conditional volatility equation allows it to capture a richer dependence structure. By computing the functional autocorrelation functions in (2.2), we assess the dependence structures of FARCH and FGARCH processes and state our first result.

Proposition 3.1. *Suppose that assumption A.1-A.3 are satisfied, the curve sequence* $\{y_i(t)\}$, $1 \leq i \leq N$, $t \in [0,1]$, *defined either in* (3.1) *or in* (3.2), *are short-range dependent in its second moment.*

In specific cases that $y_i(t)$ *follows* (3.1) *with* $p = q = 1$ *and* (3.2) *with* $\mathfrak{q} = 2$, *we have:*

first, if the sequence $\{y_i(t)\}$ *follows an FGARCH(1,1) process, the functional autocorrelation coefficient of* $\{y_i^2(t)\}$ *at lag* h *is given by*

$$\rho_h^{fgarch} = \frac{||\boldsymbol{\pi}^{h-1}\boldsymbol{\alpha} + \sum_{j=1}^{\infty}\sum_{j=h+1}^{\infty}\boldsymbol{\pi}^{2(j-1)}\boldsymbol{\alpha}^2||}{\int 1 + \sum_{j=1}^{\infty}\sum_{j=1}^{\infty}\boldsymbol{\pi}^{2(j-1)}\boldsymbol{\alpha}^2(t,t)dt},$$

where $\boldsymbol{\pi}(t,s) = \boldsymbol{\alpha}(t,s) + \boldsymbol{\beta}(t,s)$;

second, if the sequence $\{y_i(t)\}$ *follows an FARCH(2) process, the functional autocorrelation coefficient of* $\{y_i^2(t)\}$ *at lag* h *is given by*

$$\rho_h^{farch} = \frac{||\sum_{j=0}^{\infty}\sum_{j=h}^{\infty}(\sum_{k=0}^{j}\theta_1^k\theta_2^{j-k})^2||}{\int \sum_{j=0}^{\infty}\sum_{j=0}^{\infty}(\sum_{k=0}^{j}\theta_1^k\theta_2^{j-k}(t,t))^2dt},$$

where the kernel coefficients $\theta_1 + \theta_2 = \boldsymbol{\alpha}_1$ *and* $\theta_1\theta_2 = -\boldsymbol{\alpha}_2$. *In both cases, the FAC measures decay exponentially fast and converge to zero.*

Proposition 3.1 implies that both the FARCH and FGARCH models with a finite number of lags are not designed to accommodate the long-range dependence of the conditional volatility of asset returns; therefore a new model is called for.

A parsimonious model that facilitates the interpretation of long-range dependence can be obtained by introducing an exogenous covariate. Inspired by the seminal work in Engle (2002) and

others (e.g. Han, 2015), we propose a functional GARCH-X model that allows the conditional volatility $\sigma_i^2(t)$ in (3.1) to be explained by an additional non-negative covariate $\{x_i(t)\}$ (or $\{x_i^2(t)\}$), which is given by,

$$\sigma_i^2(t) = \omega(t) + \boldsymbol{\alpha}(y_{i-1}^2)(t) + \boldsymbol{\beta}(\sigma_{i-1}^2)(t) + \boldsymbol{\gamma}(x_{i-1})(t). \tag{3.3}$$

where $[\omega(t) + \boldsymbol{\gamma}(x_{i-1})(t)] \in L^2[0,1]_*^+$ for all i, and the kernel of the coefficient operators, $\alpha(t,s)$, $\beta(t,s)$ and $\gamma(t,s)$ are elements in $L^2[0,1]^+ \times L^2[0,1]^+$. The introduction of the covariate $x_i(t)$ into the model is to add richness to the dependence structure of the process, which can be modeled as being independent or dependent with the return curves. In a scalar time series context, Han (2015) discussed the properties of the GARCH-X process when the covariate X follows a stationary, fractionally integrated, or non-stationary process. His result indicates that introducing a covariate X with long-range dependence into the GARCH process can adequately explain the long memory property in volatility of daily returns. In the present paper, we do not consider non-stationary curves and focus instead only on the stationary functional curves with long-range dependence. An example of such $\{x_i(t)\}$ is that it admits a functional FARIMA(p,d,q) process specified in Li et al. (2019) that, for $d \in (0, 1/2)$,

$$\nabla^d x_i(t) = u_i(t), \quad \nabla = 1 - B, \tag{3.4}$$

where B is a backshift operator, and $u_i(t)$ is a functional ARMA(p,q) process. To suppress notations, for $p = q = 1$ we write,

$$u_i(t) - \boldsymbol{\vartheta}_1(u_{i-1})(t) = \epsilon_i(t) + \boldsymbol{\vartheta_2}(\epsilon_{i-1})(t).$$

In practice, $\{x_i(t)\}$ can be chosen from a wider range of economic variables, alternatively it can be derived from historical return data, which will be further discussed in Section 5.
We are now ready to state our next result.

Proposition 3.2. *Under assumptions A.1-A.4, there exists a stationary and non-anticipative solution to* (3.3)*; the curve sequence* $\{y_i(t)\}$*,* $1 \leq i \leq N$*,* $t \in [0,1]$*, has long-range dependence in its second moment.*

Particularly, the functional autocorrelation coefficient of $\{y_i^2(t)\}$ *at lag* h *is defined as*

$$\rho_h^{fgarchx} = \frac{||\sum_{j=0}^{\infty}\sum_{j=h}^{\infty}\boldsymbol{\pi}^{2j}\boldsymbol{\gamma}^2\boldsymbol{\varphi}_\infty + \sum_{j=1}^{\infty}\sum_{j=h+1}^{\infty}(\boldsymbol{\pi}^{(j-1)}\boldsymbol{\alpha})^2||}{\int\sum_{j=0}^{\infty}\sum_{j=0}^{\infty}\boldsymbol{\pi}^{2j}\boldsymbol{\gamma}^2(t,t)\boldsymbol{\varphi}_\infty + \sum_{j=1}^{\infty}\sum_{j=1}^{\infty}(\boldsymbol{\pi}^{(j-1)}\boldsymbol{\alpha}(t,t))^2dt},$$

where $\boldsymbol{\varphi}_\infty = \sum_{k=0}^{\infty}\sum_{k=0}^{\infty}\varphi_k^2$ *and for large* h*,*

$$h^{-2\varsigma}\rho_h^{fgarchx} < \infty, \quad as \quad h \to \infty.$$

The FAC of the FGARCH-X process decays hyperbolically with rate -2ς*.*

Proposition 3.2 guarantees the stationary solution to the FGARCH-X model, and it shows that the FGARCH-X process is long-range dependent in the second moment, providing a more natural approach to model volatility of crude oil intra-day return curves. The proofs of propositions in this section are given in Appendix A.

3.2. **Basis selections for the model estimation.**

This section focuses on estimation of the FGARCH-X model. We use the method of Cerovecki et al. (2019) and project the conditional volatility $\sigma_i^2(t)$ of (3.3) to a finite M-dimensional subspace of $L^2[0,1]^+$. The coefficients are then obtained by solving a Quasi-Likelihood optimization program. Suppose that there are M known linearly independent non-negative functions $\{\psi_1, \psi_2, \ldots, \psi_M\} \in L^2[0,1]^+$, and there exists a non-negative vector $D = [d_1, \ldots, d_M]^\top$ in $\mathbb{R}_+^M$, and non-negative matrices $A = (a_{l,m})$, $B = (b_{l,m})$, $G = (g_{l,m})$ in $\mathbb{R}_+^{M\times M}$ such that for (3.3) we have,

$$\omega = \sum_{l=1}^{M} d_l\psi_l, \quad \boldsymbol{\alpha} = \sum_{l,m=1}^{M} a_{l,m}\psi_l\psi_m, \boldsymbol{\beta} = \sum_{l,m=1}^{M} b_{l,m}\psi_l\psi_m, \boldsymbol{\gamma} = \sum_{l,m=1}^{M} g_{l,m}\psi_l\psi_m,$$

and this forms a finite parameter space,

$$\theta = vec(D, A, B, G) \in \Theta \equiv \mathbb{R}_+^{M+3\times M^2}.$$

Then, the parameter vector θ can be consistently estimated by solving

$$\hat{\theta}_N = \underset{\theta\in\boldsymbol{\Theta}}{argmin}\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{M}\{\frac{\langle y_i^2, \psi_l\rangle}{\langle\tilde{\sigma}_i^2, \psi_l\rangle} + \log\langle\tilde{\sigma}_i^2, \psi_l\rangle\}. \tag{3.5}$$

By setting initial values of $y_0(t)$ and $\sigma_0^2(t)$ as $\hat{\omega}(t)$ and unit-valued constant function, respectively, we can recursively calculate the conditional variance $\tilde{\sigma}_i^2$ via,

$$\tilde{\sigma}_i^2 = \tilde{\omega}(t) + \tilde{\boldsymbol{\alpha}}(y_{i-1}^2)(t) + \tilde{\boldsymbol{\beta}}(\sigma_{i-1}^2)(t) + \tilde{\boldsymbol{\gamma}}(x_{i-1})(t).$$

It is important to note that in the estimation we need to determine the number of basis M and the choice of basis functions ψ. Various criteria are available for selecting M, such as, the total variation explanation, which selects a big enough number of M to explain a certain proportion of the total variation; cross-validation approach; or, the diagnostic checking tests (Rice et al., 2020a). Relatively speaking, it is far more important to choose appropriate basis functions because we want that the projection scores to better capture the dependence structure of curve data, and the positivity of conditional volatilities is warranted. We note that the positivity constraint discussed here aligns with Bollerslev (1986)'s constraint, which requires that every component of conditional volatility is non-negative. We also point out that Cerovecki et al. (2019)'s heuristic data-driven approach truncates the negative part from empirical bases that are derived from $y^2(t)$ via the functional principal component analysis -which we call Truncated FPCA (TFPCA). Below we introduce two new basis functions.

3.2.1. *Functional sparse and non-negative basis (FSNN).*

A more natural way to derive data-driven basis functions is to impose a sparse and non-negative constraint to the FPCA. This method draws our interests because (i) the non-negativity can aid interpretability of results, together with the fact that the total variance explained by these non-negative bases is additive, and (ii) the sparse PCA can preserve as much as possible variation of the data under an optimal sparsity pattern to avoid a production of redundant information. In the context of multivariate time series analysis, Sigg and Buhmann (2008) developed an expectation-maximization (EM) algorithm to compute sparse non-negative empirical bases. We adapt their method to be utilised in the context of the FPCA and obtain functional basis functions for our study. Below we present the resulting algorithm,

Algorithm 1: Sparse and Non-negative PCA for Functional Data

Input: $x_i^2(t) \in L^2[0,1]$, where $i \in [0,1]$, $1 \leq i \leq N$.

Output: $\hat{\psi}_l(t)$, $l \in [1, M]$

Step 1. project $x_i^2(t)$ to a finite number K of B-spline basis functions to obtain functional loadings $\mathbf{s}_i \in \mathbb{R}^{N \times K}$, for $K \geq M$;

Step 2. apply the EM algorithm (Sigg and Buhmann, 2008) on the functional loadings $\mathbf{s}_i$ to obtain sparse and non-negative vector principal components $\hat{\epsilon}_l$, $l \in [1, M]$;

Step 3. linearly interpolate the vector principal components $\hat{\epsilon}_l$ into functional principal components $\hat{\psi}_l(t)$, $l \in [1, M]$.

In step 3, we choose the simple linear interpolation because there is only a negligible effect on the choice of interpolation method (Ramsey and Silverman, 2006).

To briefly explain the EM algorithm, we assume that the covariance matrix of $\mathbf{s_i}$ can be approximated by its first M eigenvectors. There exists a latent variable $\mathbf{z}_i$ in the principal component subspace $\mathbb{R}^M$ satisfying $\mathbb{E}(z_i) = \boldsymbol{\epsilon}_r^\top \mathbf{s}_i$, where $\boldsymbol{\epsilon}_r$ are the principal components with elements ϵ_l, $l \in [1, M]$, at the rth iteration in the optimizing process. This says that $\mathbf{s_i}$ is projected onto the rth principal component, and the principal components are estimated by solving the optimization problem,

$$\hat{\boldsymbol{\epsilon}} = \arg\min_{\boldsymbol{\epsilon}} \Big(\sum_{i=1}^{N} \mathbf{z}_i^2 \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} - 2 \sum_{i=1}^{N} \mathbf{z}_i \mathbf{s}_i^\top \boldsymbol{\epsilon} \Big).$$

The sparse and non-negative principal components can then be obtained by imposing two constraints on the optimization: $\sum_{l=1}^{M} |\epsilon|_l \leq \mathcal{B}$ and $\epsilon_l \geq 0$, $\forall l$. The upper bound $\mathcal{B}$ is selected for a desired sparsity pattern. To combine these two constraints, the intersection of the feasible regions of $\boldsymbol{\epsilon}$ is taken. The EM algorithm starts from some initial values of $\boldsymbol{\epsilon}_r$, and iteratively solves the optimization problem repeatedly over many times until the condition of $|\boldsymbol{\epsilon}_{r+1}^\top \boldsymbol{\epsilon}_r| > 1 - a$ is met, for some positive constant a.

3.2.2. *Truncated predictive factors (TPF).*

In the second approach, we resort to a predictive factor basis (Kargin and Onatski, 2008) and extend it to be used in the functional volatility models to capture the rich autocovariance structure. The main impediment in applying this technique to the functional volatility models is that its theoretical construction is originally built under a functional AR framework. However,

it is possible to overcome this hurdle by noticing that the FGARCH model can be written in an FARCH form, as shown in detail in Appendix A.3.

Following Kargin and Onatski (2008), we denote by $\mathcal{C}_0$ and $\mathcal{C}_j$, $1 \leq j \leq \mathfrak{q}$, the covariance operator and the cross-covariance operator of $y_i^2(t)$, respectively, which can be estimated by $\hat{\mathcal{C}}_0 = \frac{1}{N}\sum_{i=1}^{N}(y_i^2(t) - \bar{y}_0^2)(y_i^2(s) - \bar{y}_0^2)$ and $\hat{\mathcal{C}}_j = \frac{1}{N-j}\sum_{i=1}^{N-j}(y_i^2(t) - \bar{y}_0^2)(y_{i+j}^2(s) - \bar{y}_0^2)$, for the sample mean $\bar{y}_i^2$. We further denote $\mathfrak{R}_L$ as a set of all finite-rank operators acting on $L^2[0,1]$. There should exist a series of operators $\boldsymbol{U}_j \in \mathfrak{R}_L$ to approximate $\boldsymbol{\alpha}_j$ in (3.2) by minimizing

$$\mathbb{E}||y_i^2(t) - (\omega(t) + \boldsymbol{U}_1(y_{i-1}^2)(t) + \cdots + \boldsymbol{U}_q(x_{i-\mathfrak{q}}^2)(t))||^2.$$

Given that $\omega(t)$ is a positive constant function, this is equivalent to minimizing,

$$\mathbb{E}||(\boldsymbol{\alpha}_1 - \boldsymbol{U}_1)(y_i^2)(t) + \cdots + (\boldsymbol{\alpha}_{\mathfrak{q}} - \boldsymbol{U}_{\mathfrak{q}})(y_{i-\mathfrak{q}+1}^2)(t)||^2.$$

It is easy to see that the minimization will be attained if $E||(\boldsymbol{\alpha}_j - \boldsymbol{U}_j)(y_i^2)(t)||^2$ is minimized, for $1 \leq j \leq \mathfrak{q}$. This reduces to solving the following problem

$$\min_{U_z \in \mathcal{R}_L} ||\boldsymbol{U}_j - \boldsymbol{\alpha}_j||_{\mathfrak{s},2}, \tag{3.6}$$

where $||\cdot||_{\mathfrak{s},2}$ is a modified Hilbert-Schmidt norm. Under assumptions A.1-A.2, Equation (3.6) shares the same optimization problem with Equation (3) in Kargin and Onatski (2008). Theorem 1 in their paper indicates that (3.6) can be solved by decomposing the empirical operator $\hat{\mathbb{C}}_j = \hat{\mathcal{C}}_0^{1/2}\hat{\mathcal{C}}_j^{\top}\hat{\mathcal{C}}_j\hat{\mathcal{C}}_0^{1/2}$ which contains the information for estimating the kernel coefficients $\boldsymbol{\alpha}_j$, and the eigenfunctions decomposed from $\hat{\mathbb{C}}_j$ are expected to explain the dynamics from $x_{i-j}^2(t)$ to $x_i^2(t)$. However, the empirical eigenfunctions of $\hat{\mathbb{C}}_j$ cannot be directly used because they do not converge to the corresponding true quantities. Kargin and Onatski (2008) suggested the adoption of a regularized version of $\hat{\mathbb{C}}_{j,\tau}$ by replacing $\hat{\mathcal{C}}_0$ with $\hat{\mathcal{C}}_{0,\tau} = \hat{\mathcal{C}}_0 + \tau \cdot \mathbb{I}$, with some $\tau > 0$ and an identity matrix $\mathbb{I}$. We note that selecting the parameter τ can be tricky for the estimation. Here, we choose $\tau = 0.75$ by drawing on the past empirical experience. Eventually, the data-driven PF $\hat{f}_{j,l}$ is obtained as,

$$\hat{f}_{j,l}(t) = \left\langle x_i(t), \hat{\mathcal{C}}_{0,\tau}^{-1/2}\hat{v}_{j,l}(t)\right\rangle, \quad 1 \leq l \leq M,$$

where $\hat{v}_{j,l}(t)$ is the eigenfunctions of $\hat{\mathbb{C}}_{j,\tau}$. Meanwhile, for the sake of securing the positivity constraint, we apply a truncation trick to eliminate their negative parts to yield,

$$\hat{f}_{j,l}(t) = \inf_{t\in[0,1]} \hat{f}_{j,l}(t) \wedge 0, \text{ for all } l = 1, \ldots, M.$$

4. A simulation study to compare functional volatility processes

In this section we consider three functional volatility processes, viz, FGARCH(1,1), FARCH(5), and FGARCH-X, and conduct a simulation study to assess their dependence structures. Choosing lags up to 5 for the FARCH model is of particular interest given the inclusion of information at a weekly level in practice. The data are generated through the mean equation of (3.1), and their volatility processes, which follow (3.1), (3.2), and (3.3), respectively. We use an Ornstein-Uhlenbeck process as the innovation curve in the mean equation,

$$\varepsilon_i(t) = e^{-t/2} W_i(e^t), t \in [0, 1], \tag{4.1}$$

where $W_i(\cdot)$ are iid standard Brownian motions. Then, the kernel coefficients in the volatility equations are specified respectively as:

$$\omega(t) = \chi_0 t(1-t), \quad \boldsymbol{\alpha}(t,s) = \chi_1 t(1-t)s(1-s)$$

$$\boldsymbol{\beta}(t,s) = \chi_2 t(1-t)s(1-s), \quad \boldsymbol{\gamma}(t,s) = \chi_3 t(1-t)s(1-s),$$

where the values of parameters $\chi_{0,1,2,3}$ are suggested by the empirical data and determined in Table 4.1. The covariate $\{x_i(t)\}$ in the FGARCH-X process is assumed to follow a functional FARIMA(1,d,1) process, c.f. (3.4), where the kernel coefficients

$$\boldsymbol{\vartheta}_1(t,s) = 0.34 \exp(-(t^2+s^2)/2), \quad \boldsymbol{\vartheta}_2(t,s) = \frac{3}{2}\min(t,s).$$

Here we specify $d = 0.45$ so that $x_i(t)$ is a weakly stationary curve sequence with relatively strong long-range dependence. Each sample contains 500 observation with a grid point of 50. These processes are generated with 1000 replications, and the FAC up to lag 50 is calculated correspondingly.

TABLE 4.1. The selected values of the parameters χ_0, χ_1, χ_2, and χ_3.

	χ_0	χ_1					χ_2	χ_3
		y_{i-1}^2	y_{i-2}^2	y_{i-3}^2	y_{i-4}^2	y_{i-5}^2	σ_{i-1}^2	x_{i-1}^2
FGARCH(1,1)	0.5	6	-	-	-	-	16	-
FARCH(5)	0.5	8	4	3	3	3	-	-
FGARCHX	0.5	6	-	-	-	-	16	8

Figures 4.1 - 4.3 display the FAC plots of the squared simulated return data. The solid lines indicate the averaged FAC over 1000 replications, while the dotted lines represent their confidence intervals at the 95% significance level. In the case of the FGARCH(1,1) process, the FAC of the squared process decays exponentially fast to an iid level; a similar pattern is also observed for the squared FARCH(5) in Figure 4.2 when the lag exceeds a certain number around 5. These results are in concordance with the property of short-range dependence stated in Proposition 3.1. However, the FAC of the squared FGARCH-X in Figure 4.3 exhibits a pattern more similar to the autocorrelation of real data as displayed in Figure 2.2. In this case, the FAC decreases exponentially fast at few initial lags, and then shows a hyperbolic decay for a large number of lags. In summary, this result numerically confirms the claim made in the second part of Proposition 3.2.

FIGURE 4.1. FAC Plots of the squared returns of simulated FGARCH(1,1) process.

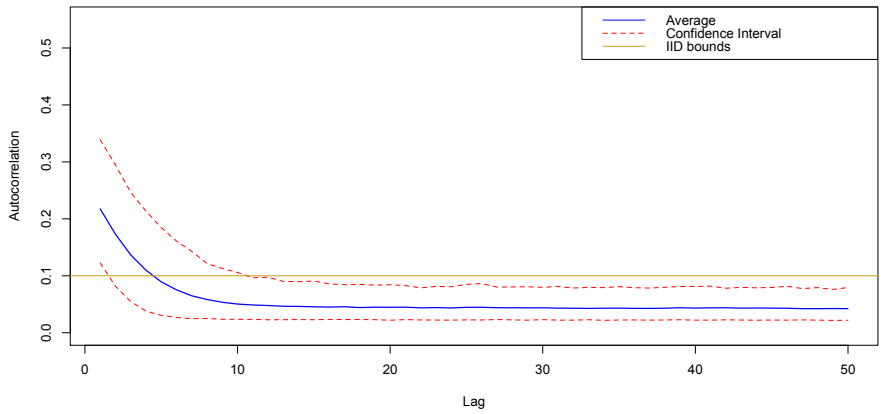


FIGURE 4.2. FAC Plots of the squared returns of simulated FARCH(5) process.

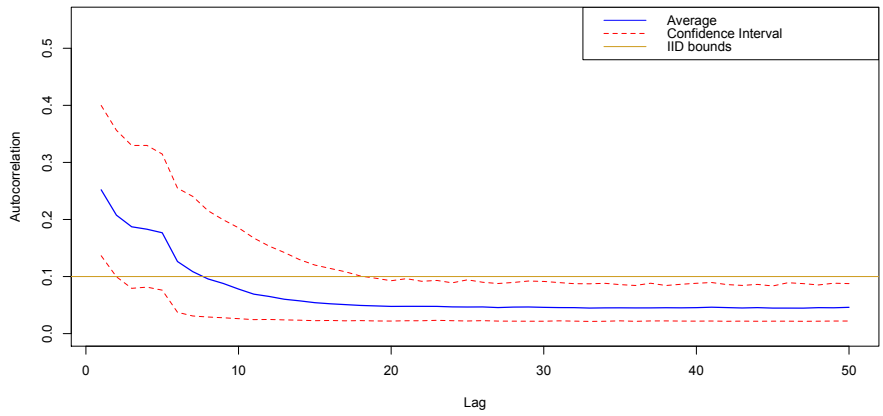
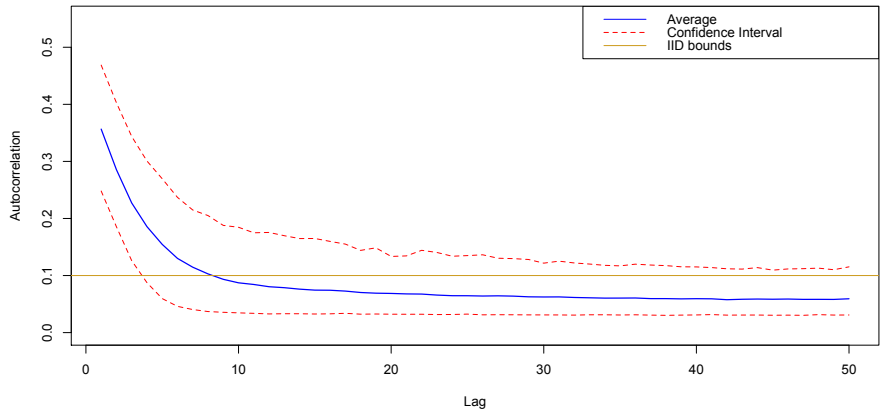


FIGURE 4.3. FAC Plots of the squared returns of simulated FGARCH-X process.



5. MODELING AND FORECASTING THE VOLATILITY OF CRUDE OIL RETURN CURVES

In this section, we empirically model the volatility of crude oil intra-day return curves and perform a one-day-ahead forecast. The out-of-sample exercise is conducted between the beginning of 2017 and the beginning of May 2020, allowing each training sample to fit 500 curves. To this end, we re-estimate the models with a rolling window approach for every quarter (63 trading days) to accommodate new information from evolving market conditions. We remark that all of the models discussed in the paper are able to provide multistep-ahead forecasts, with a potentially substantial increase in forecasting errors and computational requirement.

5.1. **Model specification and in-sample fitting.**

To be consistent with the simulation study in Section 4, the FGARCH(1,1), FARCH(5) and

the FGARCH-X models are considered as model candidates. We implement the FGARCH-X model, and the list of the covariate X chosen in the estimation of the FGARCH-X model is as follows:

(1) **HMV** – historical mean variation curve

We first construct the covariate X by averaging the absolute value of historical intra-day log returns:

$$x_i(t) = \frac{1}{m}\sum_{j=1}^{m} |r_{i-j}(t)|, \tag{5.1}$$

where $r_i(t) = 100[\log P_i(t) - \log P_i(t-\Delta)]$, $\Delta = 5min$. We set $m = 5$ given that the trading information within one week is included. This estimator is inspired by Corsi (2009) who averaged past realized volatilities to model the long-memory property of daily returns. It is important to point out that equation (5.1) is not a realized estimator although it contains the same information measured as the realized volatility.

(2) **IRV** – intra-day range variation curve

In the context of the scalar volatility modeling, Christensen and Podolskij (2007) introduced a realized range estimator that is more efficient than realized volatility in a market frictionless world. Here we introduce a curve-type intra-day range estimator to preserve this information:

$$x_i(t) = 100|\log P_i^h(t) - \log P_i^l(t)|$$

where $P_i^h(t)$ and $P_i^l(t)$ are respectively the maximum and minimum prices of 5-minute intervals.

(3) **IBV** – intra-day bipower variation curve

Another proxy variable is enlightened following the discussion in Barndorff-Nielsen and Shephard (2006), who defined a realized bipower variation estimator that is immune to jumps. Although we do not consider jumps in the context of the functional data analysis, a similar curve estimator can be assembled to depict variations in the curve:

$$x_i(t) = |r_i(t)||r_i(t-\Delta)|$$

where intra-day log return $r_i(t)$ is defined in Equation (5.1).

(4) **IVOL** – intra-day volume curve

Lastly, we use an intra-day volume curve, and we refer to Fuertes et al. (2009) for discussions about the performance of trading volume in the context of the scalar GARCH-X model. Since the trading volumes from intra-day intervals $VOL_i(t)$ form a non-stationary process, we proceed with the usual log transformation:

$$x_i(t) = |\log VOL_i(t) - \log VOL_i(t - \Delta)|.$$

To further understand the properties of the covariate processes, we estimate the memory parameter "LRS-d" (Li et al., 2019) and calculate the resulting statistic summaries. The properties are summarized by using the first training sample (500 observations from January 2015 to December 2016). The same procedure could be recursively performed for each rolling window of the training sample. To save some computational costs, we skip this repetitive analysis, and in an unreported analysis, we find that the results are generally consistent with randomly selected training samples. Table 5.1 shows that the OCIDR curves themselves are short-range dependent with LRS-d $= 0.04$. Also, we find that the chosen four X covariates are weakly stationary, non-normal, autocorrelated, and conditionally heteroscedastic. The LRS-d parameters range between 0.23 and 0.29, indicating a property of long-range dependence. This leads us to conclude that the covariates $x_i(t)$ discussed above are suitable for the FGARCH-X model in order to explain the persistence in the second moment of crude oil OCIDR curves.

TABLE 5.1. Statistical summary of the OCIDR and X covariates with the P-values documented for stationary, normality, autocorrelation, and conditional heteroscedasticity tests.

	No.Obs	Stationary	Normality	Autocorrelation				Heteroscedasticity				LRS-d
H				1	5	10	20	1	5	10	20	
OCIDR	500	0.31	0.00	0.99	0.34	0.75	0.17	0.00	0.00	0.00	0.00	0.04
HMV	500	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29
IRV	500	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27
IBV	500	0.24	0.00	0.00	0.01	0.00	0.03	0.00	0.00	0.00	0.00	0.23
IVOL	500	0.16	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.23

We now assess the in-sample fitting of the model candidates. As discussed in section 3.2, choosing M data-driven bases is essential for the dimension reduction in deciding for the model estimation. Applying the TFPCA, FSNN, and TPF methods, we find that the first two bases from all of the approaches used in the study account for over 90% of the total variation, thereby

prompting us to set $M = 2$ and treating the remaining bases as noises. Note the selected basis functions are normalized with a unit norm. Figure 5.1 displays the estimation result for the kernel coefficients $\alpha(t, s)$, $\beta(t, s)$ and $\gamma(t, s)$ in Equation (3.3) when HMV is used as the covariate X. The first and second rows represent the estimated kernel operators with FSNN and TPF basis functions. The results of the TFPCA bases are suppressed because their values are very close to the FSNN after the normalization. From the plot, we note that the magnitude of the estimated operators $\beta(t, s)$ and $\gamma(t, s)$ are much larger than the ARCH-effect operator $\alpha(t, s)$, indicating that the crude oil intra-day return curves exhibit a strong persistence volatility effect.

FIGURE 5.1. Plots of the estimated kernel functions in the FGARCH-X(HMV) model with the first row presenting the estimators using the FSNN bases and the second row presenting the estimators using the TPF bases for $M = 2$.

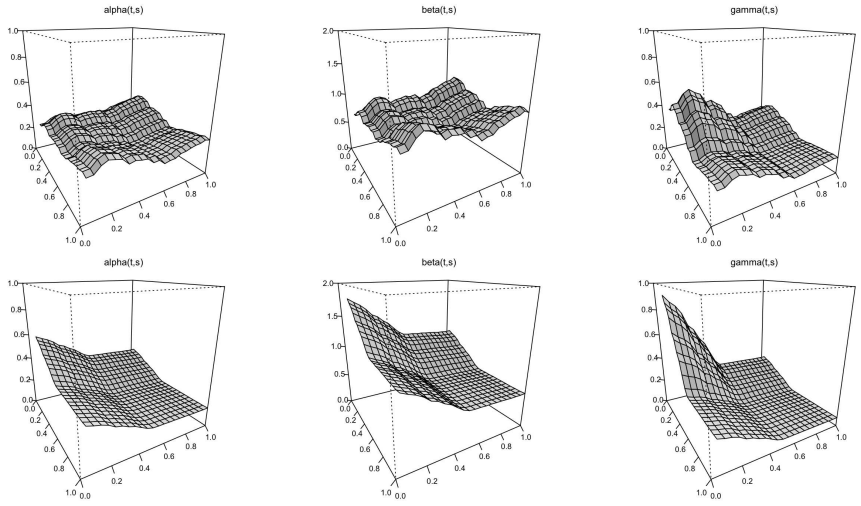


Table 5.2 reports the norm of estimated coefficients of each model. The columns D, A_{1-5}, B and G represent the non-negative coefficient matrices used to obtain operators $\omega(t)$, $\alpha(t, s)$, $\beta(t, s)$, and $\gamma(t, s)$, respectively, and the calculated Euclidean norms show persistence of the conditional volatility process. The overall results are in line with Figure 5.1. The TFPCA and FSNN bases produce the same performances as similar information is being captured. Besides, Table 5.2 also reports the P-values of the goodness-of-fit test for each of the model candidates. We apply the $M_{N,K,\varepsilon}$ test by Rice et al. (2020a) to diagnostically check dependence of model fitted errors at lags $\{1, 5, 10, 20\}$. We find that the proposed FGARCH-X models are able to fit conditional heteroscedasticity in the return curves adequately, in particular when the TPF bases are adopted. Relatively speaking, the existing FGARCH-type models are less adequate

in this dimension, when they are tested at lag $H = 20$; and lastly the FARCH-type models are generally much less adequate at fitting conditional heteroscedasticity at higher lags.

Furthermore, we also plot the FAC of squared curves from the fitted FGARCH-X model. Figure 5.2 displays the FAC plot of the squared fitted curves from these models when the covariate X is chosen as HMV, IRV, IBV, and IVOL, respectively. Comparing with the counterparts obtained from the true squared observations as well as the fitted FARCH and FGARCH models as shown in Figure 2.2, it is noticeable that the FGARCH-X models are generally more suitable for capturing long-range dependence and are able to restore the empirical FAC.

TABLE 5.2. The Euclidean norm of estimated coefficient matrices and the P-values of goodness-of-fit test on model candidates.

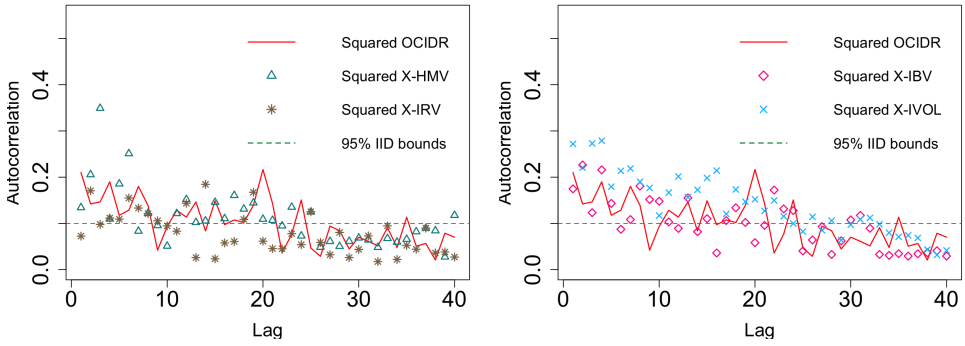
		Kernel estimation								Goodness-of-fit			
		D	A_1	A_2	A_3	A_4	A_5	B	G	H=1	H=5	H=10	H=20
FGARCH(1,1)	TFPCA	0.10	0.22	-	-	-	-	0.75	-	0.63	0.32	0.65	0.04
	FSNN	0.10	0.22	-	-	-	-	0.75	-	0.63	0.32	0.65	0.04
	TPF	0.11	0.15	-	-	-	-	0.47		0.26	0.23	0.38	0.01
FARCH(5)	TFPCA	0.39	0.21	0.04	0.06	0.07	0.08	-	-	0.15	0.02	0.00	0.00
	FSNN	0.39	0.21	0.03	0.06	0.07	0.09	-	-	0.14	0.01	0.00	0.00
	TPF	0.42	0.26	0.02	0.06	0.03	0.08	-	-	0.21	0.00	0.00	0.00
FGARCH-X (HMV)	TFPCA	0.55	0.13	-	-	-	-	0.95	0.92	0.72	0.94	0.91	0.07
	FSNN	0.56	0.13	-	-	-	-	0.95	0.94	0.72	0.95	0.91	0.07
	TPF	0.46	0.12	-	-	-	-	0.44	0.95	0.73	0.98	0.75	0.22
FGARCH-X (IRV)	TFPCA	0.34	0.10	-	-	-	-	0.95	0.90	0.76	0.96	0.92	0.07
	FSNN	0.34	0.10	-	-	-	-	0.95	0.90	0.76	0.96	0.92	0.07
	TPF	0.11	0.08	-	-	-	-	0.42	0.94	0.75	0.98	0.79	0.25
FGARCH-X (IBV)	TFPCA	0.85	0.14	-	-	-	-	0.96	0.95	0.69	0.93	0.93	0.04
	FSNN	0.83	0.14	-	-	-	-	0.84	0.94	0.68	0.92	0.93	0.03
	TPF	0.76	0.12	-	-	-	-	0.47	0.95	0.73	0.97	0.79	0.12
FGARCH-X (IVOL)	TFPCA	0.89	0.16	-	-	-	-	0.95	0.04	0.62	0.90	0.91	0.02
	FSNN	0.88	0.16	-	-	-	-	0.58	0.02	0.60	0.89	0.91	0.02
	TPF	0.83	0.15	-	-	-	-	0.51	0.04	0.72	0.96	0.76	0.08

5.2. **Forecasting Value-at-Risk for intra-day return curves.**

To evaluate the out-of-sample performance, we compute one-day-ahead Value-at-Risk (VaR) by using the predicted conditional volatilities. From the mean equation in (3.1), the intra-day $\widehat{\mathrm{VaR}}^{\tau}(t)$ at day $i+1$ can be calculated as:

$$\widehat{\mathrm{VaR}}^{\tau}_{i+1}(\cdot) = \widehat{\sigma}_{i+1}(\cdot)\widehat{\varepsilon}^{\tau}(\cdot), \quad t \in [0,1] \tag{5.2}$$

FIGURE 5.2. FAC plots of the squared OCIDR curves and the squares of FGARCH-X model fitted curves with the FSNN basis.



where $\hat{\sigma}^2_{i+1}(t)$ is the predicted intra-day volatility curve obtained from the functional volatility models, and $\hat{\varepsilon}^{\tau}(t)$ is the unconditional quantile of the error process at a selected significance level τ. To find suitable estimators for $\hat{\varepsilon}^{\tau}(t)$, we consider three types of error processes: (i) a Gaussian process, i.e., the series at each time point t follows a standard normal distribution; (ii) a process with observations at each time point t following a student-t distribution with 5 degrees of freedom; and (iii) an empirical process obtained from iid bootstrapping the residual curves of the FGARCH, FARCH, and FGARCH-X models with 1000 replications.

These intra-day VaR curves provide a valid evaluation on volatility forecasting as the evolution of intra-day return curves should not exceed the VaR curves at a certain significance level if the volatility forecasts are sufficiently accurate. This can be measured by counting the number of times that intra-day return curves cross the VaR curves, i.e., the number of the violations. Following Rice et al. (2020b), the point-wise violation process $Z^{\tau}_i(t)$ is defined as,

$$Z^{\tau}_i(\cdot) = \mathbb{1}_{\{y_i(\cdot) < \widehat{\mathrm{VaR}}^{\tau}_i(\cdot)\}}, \text{ for any } t \in [0,1].$$

where $\mathbb{1}$ is an indicator function and the process $Z^{\tau}_i(t)$ is composed of values between 0 and 1, representing no exceedance and point-wise exceedance over the intra-day interval, respectively. These violation curves should be unbiased and independent if the intra-day VaR curves are valid. Importantly, we apply the unbiasedness and independence tests (Rice et al., 2020b) to backtest the following hypothesis on the violation curves:

$$\mathcal{H}^{(1)}_0\text{: } \mathbb{E}[Z^{\tau}_i(\cdot) - \tau]\text{=0} \quad \text{and} \quad \mathcal{H}^{(2)}_0\text{: } Z^{\tau}_i(\cdot) \text{ is IID along } i.$$

Table 5.3 displays the P-values of backtests for the intra-day VaR curves forecasts at the nominal levels $\tau = 0.05$ and 0.01. Panel A reports the results of the unbiasedness test, which indicate that

all of the model candidates, except for the FGARCH-X (IVOL) model with TFPCA and FSNN bases, produce unbiased violations to the nominal levels, even when we use the unconditional quantile taken from the Gaussian-type errors. Furthermore, Panel B shows the P-values of independence test with a maximum lag length $H \in \{1, 5, 10\}$. Overall, all of the models perform reasonably in the independent backtesting. The null hypothesis is rejected in a few cases of the FGARCH(1,1) and FGARCH-X models, and this is particularly manifested when we use the TPF basis functions and the Gaussian or Student-T errors are employed. Notably, the FGARCH-X (HMV) model passes all the backtests regardless of the basis function selection. Besides, Figure 5.3 plots the P-values of the independence tests on the violation processes with $\tau = 0.01$ obtained from the model candidates by using the TPF basis and bootstrapped errors over the lag length from $H = 1$ up to 20. In a nutshell, despite promising performances, the FGARCH-X models do not appear to outperform the other models unambiguously, and the FGARCH(1,1) model is still competitive enough to forecast valid intra-day VaR curves. A similar finding is also reported in the scalar context that the simple GARCH(1,1) model is not outperformed by more sophisticated in-sample fitting models (Hansen and Lunde, 2005). Explaining other potential stylized features, such as the leverage effect (Sun and Yu, 2020), may improve the out-of-sample forecasting of volatility, but this is beyond the main scope of this article, so we leave it as future research work.

FIGURE 5.3. P-values of the independence backtest for the FARCH(1), FGARCH(1,1) and FGARCH-X models applied to the violation processes with 1% nominal level at a function of H, $H = 1, \ldots, 20$.

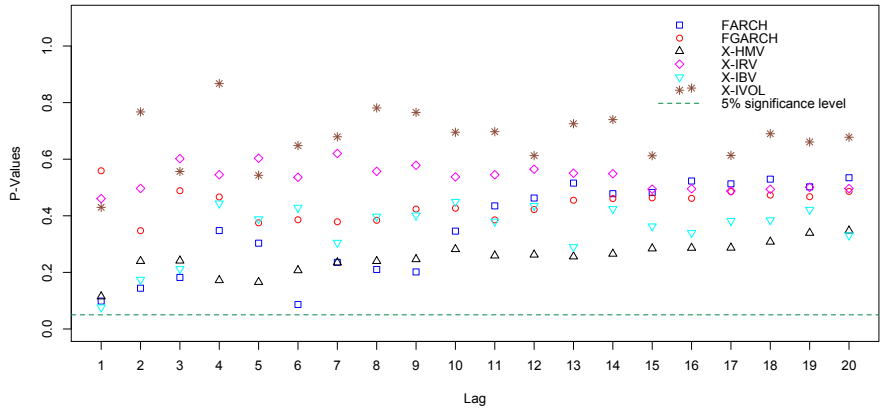


TABLE 5.3. The P-value of backtests of the unbiasedness and the independence for the VaR curves forecasts with the bold values indicating the significance at the 5% Level.

		FGARCH(1,1)			FARCH(5)			X-HMV			X-IRV			X-IBV			X-IVOL		
		TFPCA	FSNN	TPF	TFPCA	FSNN	TPF	TFPCA	FSNN	TPF	TFPCA	FSNN	TPF	TFPCA	FSNN	TPF	TFPCA	FSNN	TPF
Panel A: $\mathcal{H}_0^{(1)}$: the unbiasedness test																			
Gaussian	$\tau = 0.05$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.33	0.34	1.00	**0.00**	**0.00**	**0.00**
	$\tau = 0.01$	0.99	1.00	0.91	0.86	0.87	0.72	1.00	1.00	0.99	1.00	1.00	0.79	1.00	1.00	1.00	0.12	0.12	0.34
Student-T	$\tau = 0.05$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.24	0.23	1.00	**0.00**	**0.00**	**0.00**
	$\tau = 0.01$	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	0.35	0.34	1.00	**0.00**	**0.00**	**0.00**
Bootstrap	$\tau = 0.05$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.91	0.86	1.00	**0.00**	**0.00**	**0.00**
	$\tau = 0.01$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.13	0.09	0.09
Panel B: $\mathcal{H}_0^{(2)}$: the independence test, $H = 1$																			
Gaussian	$\tau = 0.05$	0.54	0.51	0.45	0.60	0.54	0.36	0.65	0.57	0.68	0.40	0.67	0.78	0.59	0.17	0.56	0.36	0.36	0.08
	$\tau = 0.01$	0.59	0.55	0.21	0.54	0.66	0.38	0.37	0.40	0.62	0.55	0.57	0.37	**0.03**	**0.02**	0.50	0.53	0.50	0.51
Student-T	$\tau = 0.05$	0.49	0.44	0.47	0.53	0.57	0.38	0.50	0.50	0.65	0.45	0.63	0.97	0.45	0.40	0.51	0.44	0.44	0.19
	$\tau = 0.01$	0.27	0.36	0.11	0.51	0.42	0.30	0.50	0.42	0.39	0.43	0.34	0.23	0.45	0.45	0.07	**0.03**	**0.03**	**0.00**
Bootstrap	$\tau = 0.05$	0.51	0.52	0.65	0.74	0.67	0.63	0.54	0.58	0.60	0.44	0.55	0.32	0.56	0.32	0.38	**0.00**	**0.02**	0.24
	$\tau = 0.01$	0.45	0.41	0.35	0.00	0.10	0.27	0.20	0.19	0.67	0.54	0.45	0.24	**0.00**	0.07	0.06	1.00	0.80	0.79
$\mathcal{H}_0^{(2)}$: the independence test, $H = 5$																			
Gaussian	$\tau = 0.05$	0.10	0.15	**0.01**	0.58	0.57	0.21	0.14	0.08	0.06	0.14	0.20	0.03	0.46	0.39	**0.05**	0.46	0.32	0.29
	$\tau = 0.01$	0.42	0.41	**0.04**	0.45	0.46	0.17	0.35	0.33	0.11	0.45	0.47	0.11	0.45	0.76	0.69	0.42	0.59	0.68
Student-T	$\tau = 0.05$	0.14	0.09	**0.01**	0.56	0.60	0.26	0.08	0.05	0.07	0.20	0.25	**0.03**	0.50	0.52	**0.02**	0.32	0.93	0.15
	$\tau = 0.01$	0.37	0.37	**0.02**	0.38	0.38	**0.02**	0.46	0.43	0.38	0.33	**0.04**	0.29	0.44	0.44	0.39	1.00	0.49	0.27
Bootstrap	$\tau = 0.05$	0.37	0.24	0.18	0.42	0.37	0.57	0.09	0.06	0.19	0.07	0.12	**0.05**	0.29	0.23	0.24	0.44	0.55	0.48
	$\tau = 0.01$	0.36	0.38	0.23	0.29	0.30	0.35	0.18	0.19	0.45	0.56	0.59	0.26	0.37	0.40	0.16	0.64	0.66	0.91
$\mathcal{H}_0^{(2)}$: the independence test, $H = 10$																			
Gaussian	$\tau = 0.05$	0.14	0.17	**0.02**	0.59	0.58	0.38	0.18	0.18	0.15	0.33	0.38	0.07	0.55	0.53	0.11	0.35	0.13	0.38
	$\tau = 0.01$	0.47	0.44	**0.00**	0.49	0.48	0.20	0.42	0.41	0.23	0.48	0.48	0.19	0.39	0.41	0.75	0.60	0.64	0.06
Student-T	$\tau = 0.05$	0.17	0.11	**0.01**	0.59	0.59	0.40	0.16	0.17	0.19	0.36	0.39	**0.04**	0.54	0.52	0.18	0.33	0.31	0.14
	$\tau = 0.01$	0.43	0.40	**0.00**	0.46	0.46	**0.03**	0.40	0.44	0.56	0.41	0.44	0.40	0.43	0.42	0.35	0.90	0.69	0.34
Bootstrap	$\tau = 0.05$	0.40	0.27	0.32	0.57	0.51	0.67	0.14	0.19	0.29	0.21	0.18	0.06	0.48	0.43	0.36	0.48	0.43	0.36
	$\tau = 0.01$	0.39	0.40	0.32	0.37	0.34	0.35	0.26	0.28	0.44	0.53	0.51	0.29	0.41	0.37	0.28	0.63	0.64	0.71

6. CONCLUSION

This paper considers the topic of modeling volatility of crude oil intra-day return curves. We find that the overnight cumulative intra-day return curves derived from the WTI crude oil commodity futures are serially uncorrelated and conditionally long-range heteroscedastic. In order to fit this stylized feature, we propose a functional GARCH-X model given the fact that the existing functional ARCH and GARCH-typed models only capture the short range conditional heteroscedasticity. The functional autocorrelation coefficients of the squared processes are derived to allow us to study the dependence structure of the functional ARCH, GARCH and GARCH-X models. Also, we introduce two types of bases that ensure the estimation process to generate non-negative volatility curves and produce more accurate out-of-sample forecasts, including the functional sparse and non-negative basis, and truncated predictive factor. A set of simulation study confirms that an FGARCH-X process is long-range dependent on its second moment if the covariate X is specified to be long-range dependent.

Empirically, we consider four exogenous covariates derived from intra-day returns and volumes to capture the long-range dependence in crude oil intra-day return curves. Our results show that the FGARCH-X model provides relatively more reliable in-sample fitting performances to explain the long-range dependence property. In terms of out-of-sample forecasting, all

functional volatility models are useful at producing Value-at-Risk curves for the purpose of conducting an intra-day risk management exercise, however the in-sample out-performance of the FGARCH-X model does not appear carry over to a marked improvement in terms of the out-of-sample forecast relative to the standard FGARCH(1,1) model. As avenues for future work, we may explore and model further potential features of asset intra-day return curves and improve the model predictability. Also, we may focus on exploring the impacts of potentially non-stationary exogenous covariates on the FGARCH-X models and their forecasting performances.

Appendix A. Proofs of the main results

A.1. **Proof of Proposition 3.1.**

To obtain the functional volatility models, the following regularity conditions are assumed.

Assumption A.1. *The sequence* $\{\varepsilon_i\}$ *is independent and identically distributed (iid) with zero mean* $\mathbb{E}[\varepsilon_i|\mathcal{F}_{t-1}] = 0$ *and positive definite covariance function* $c_\varepsilon(t,s) = \mathbb{E}(\varepsilon_0(t)\varepsilon_0(s))$*, and* $\mathbb{E}[\varepsilon_i^2(t)|\mathcal{F}_{t-1}] = 1$.

$\mathcal{F}_t$ is the filtration containing information available at time t. The condition $\mathbb{E}[\varepsilon_i^2(t)] = 1$ ensures the model identifiability of 3.1.

Assumption A.2. $\lim_{i\to\infty} \frac{1}{i} \log ||\prod_{j=0}^{\infty} \boldsymbol{\Psi}_{i-j}|| < 0$*, where the operator* $\boldsymbol{\Psi}_i \in \mathcal{H}^{p+q}$ *is defined as,*

$$\boldsymbol{\Psi}_i = \begin{pmatrix} \boldsymbol{\alpha}_1(\varepsilon_i^2) & \dots & \boldsymbol{\alpha}_{q-1}(\varepsilon_i^2) & \boldsymbol{\alpha}_q(\varepsilon_i^2) & \boldsymbol{\beta}_1(\varepsilon_i^2) & \dots & \boldsymbol{\beta}_{p-1}(\varepsilon_i^2) & \boldsymbol{\beta}_p(\varepsilon_i^2) \\ \boldsymbol{I}_{\mathcal{H}} & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \dots & \boldsymbol{I}_{\mathcal{H}} & 0 & 0 & \dots & 0 & 0 \\ \boldsymbol{\alpha}_1 & \dots & \boldsymbol{\alpha}_{q-1} & \boldsymbol{\alpha}_q & \boldsymbol{\beta}_1 & \dots & \boldsymbol{\beta}_{p-1} & \boldsymbol{\beta}_p \\ 0 & \dots & 0 & 0 & \boldsymbol{I}_{\mathcal{H}} & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & \boldsymbol{I}_{\mathcal{H}} & 0 \end{pmatrix}.$$

Assumption A.2 is a condition on the top Lyapunov exponent of the FGARCH(p,q) model, ensuring a unique strictly stationary and non-anticipative solution. As the main object of interest is to study the squared process, we also generally assume that a higher moment of $y_i(t)$ exists throughout the paper.

Assumption A.3. $E||y_i||^r < \infty$, *for some* $r \geq 8$, *and* $i \in \mathbb{Z}$.

Following Li et al. (2019), we define the term of long range dependence as follows.

Definition A.1. *Consider a curve sequence* $x_i(t)$ *that follows a linear process represented as an* $MA(\infty)$,

$$x_i(t) = \sum_{j=0}^{\infty} \boldsymbol{\varphi}_j(v_{i-j})(t), \tag{A.1}$$

where $\{v_i(t)\}$ *is an iid error sequence, and the integral operators are equipped with the kernel coefficients* $\varphi_j(t,s)$ *such that* $\boldsymbol{\varphi}_j(v)(t) = \int \varphi_j(t,s)v(s)ds$, $v \in L^2[0,1]$. *We say that* $x_i(t)$ *is a short-range dependent curve if the supreme of* L^2 *norm* $\{\sup ||\boldsymbol{\varphi}_j||\}$, $j > 0$ *is summable;* $x_i(t)$ *said to be long-range dependent if the property of summability is not satisfied.*

We next prove Proposition 3.1 by taking $p = q = 1$ in (3.1) and $\mathfrak{q} = 2$ in (3.2) without loss of generality. Denoting the iid sequence $\eta_i(t) = y_i^2(t) - \sigma_i^2(t)$ with zero mean and unit variance, (3.1) and (3.2) can be written as a functional ARMA representation for the squared process. According to Spangenberg (2013), there exists a functional linear process for each of the functional ARMA-typed process. Thus, $\{y_i^2\}$ from an FGARCH(1,1) follows,

$$\begin{aligned} y_i^2 &= \omega + \boldsymbol{\pi}(y_{i-1}^2) + \eta_i - \boldsymbol{\beta}(\eta_{i-1}) \\ &= \sum_{j=0}^{\infty} \boldsymbol{\pi}^j \omega + \eta_i + \sum_{j=1}^{\infty} \boldsymbol{\pi}^{j-1}\boldsymbol{\alpha}\eta_{i-j}, \end{aligned} \tag{A.2}$$

where $\boldsymbol{\pi} = (\boldsymbol{\alpha} + \boldsymbol{\beta})$. Equation (A.2) takes the form of (A.1) in Definition A.1. The short-range dependence structure of the FGARCH(1,1) model can be verified by showing the summability of $\sup ||\boldsymbol{\pi}^{j-1}\boldsymbol{\alpha}||$. Under assumption A.2, Cerovecki et al. (2019) in their Proposition 1 showed that for some $i \geq 1$, $\mathbb{E}\log ||(\boldsymbol{\pi}_{i-1} \ldots \boldsymbol{\pi}_1)|| < 0$ in the case of $p = q = 1$, where $\boldsymbol{\pi}_i = \boldsymbol{\alpha}\varepsilon_i^2 + \boldsymbol{\beta}$. Recall $\mathbb{E}\varepsilon_i^2 = 1$, this leads $||\boldsymbol{\pi}|| < 1$ and $||\boldsymbol{\pi}^j||$ being summable, by using Cauchy-Schwarz inequity,

$$\sum_{j=1}^{\infty} \sup ||\boldsymbol{\pi}^{j-1}\boldsymbol{\alpha}|| \leq \sum_{j=1}^{\infty} \sup ||\boldsymbol{\pi}^{j-1}|| \sup ||\boldsymbol{\alpha}|| \leq \sum_{j=1}^{\infty} \sup ||\boldsymbol{\pi}^{j-1}|| = C_1 < \infty,$$

where C_1 is some positive finite number.

We further calculate the autocovariance operator $\lambda_h(t,s)$ to obtain the functional autocorrelation coefficient ρ_h of the FGARCH(1,1) model at lag h, $h \geq 1$. By substituting (A.2),

$$\begin{aligned}
\lambda_h &= \mathbb{E}[(y_{i-h}^2 - \mathbb{E}(y^2))(y_i^2 - \mathbb{E}(y^2))] \\
&= \mathbb{E}[(\eta_{i-h} + \sum_{j=1}^{\infty} \boldsymbol{\pi}^{j-1}\boldsymbol{\alpha}\eta_{i-h-j})(\eta_i + \sum_{j=1}^{\infty} \boldsymbol{\pi}^{j-1}\boldsymbol{\alpha}\eta_{i-j})] \\
&= \mathbb{E}[\eta_{i-h}\eta_i + \eta_{i-h}\sum_{j=1}^{\infty} \boldsymbol{\pi}^{j-1}\boldsymbol{\alpha}\eta_{i-j} + \eta_i \sum_{j=1}^{\infty} \boldsymbol{\pi}^{j-1}\boldsymbol{\alpha}\eta_{i-h-j} + \sum_{j=1}^{\infty} \boldsymbol{\pi}^{j-1}\boldsymbol{\alpha}\eta_{i-h-j}\sum_{j=1}^{\infty} \boldsymbol{\pi}^{j-1}\boldsymbol{\alpha}\eta_{i-j}] \\
&= \boldsymbol{\pi}^{h-1}\boldsymbol{\alpha} + \sum_{j=1}^{\infty}\sum_{j=h+1}^{\infty} \boldsymbol{\pi}^{2(j-1)}\boldsymbol{\alpha}^2 \leq \sum_{j=1}^{\infty}\sum_{j=h}^{\infty} \boldsymbol{\pi}^{2(j-1)}\boldsymbol{\alpha}^2 \leq C_2 < \infty,
\end{aligned}$$

for some positive finite number C_2. The upper bound tells us that the FAC of the FGARCH(1,1) process decays exponentially fast. When $h = 0$, we have the variance term $\lambda_0 = 1 + \sum_{j=1}^{\infty} \boldsymbol{\pi}^{2(j-1)}\boldsymbol{\alpha}^2$, and ρ_h^{fgarch} is obtained.

Next, we consider the FARCH(2) process with a squared process $\{y_i^2\}$ having the form of a linear process,

$$\begin{aligned}
y_i^2 &= \sum_{j=0}^{\infty}(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2)^j\omega + (1-\theta_1 B)^{-1}(1-\theta_2 B)^{-1}\eta_i \\
&= \sum_{j=0}^{\infty}(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2)^j\omega + \sum_{j=0}^{\infty}(\sum_{k=0}^{j}\theta_1^k\theta_2^{j-k})(\eta_{i-j}),
\end{aligned} \tag{A.3}$$

where B is a backshift operator, and $\theta_1 + \theta_2 = \boldsymbol{\alpha}_1$, $\theta_1\theta_2 = -\boldsymbol{\alpha}_2$. It is not hard to see that $\sum_{j=0}^{\infty}(\sum_{k=0}^{j}\theta_1^k\theta_2^{j-k})(\eta_{i-j}) < \sum_{j=0}^{\infty}(\theta_1+\theta_2)^j(\eta_{i-j})$. Given that $(\theta_1+\theta_2) < (\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)$, we therefore have the summability of $\sup||(\theta_1+\theta_2)^j||$ if $\sup||(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)^j||$ is summable. Under assumption A.2, we deduce that for some $i \geq 1$, $\mathbb{E}\log||(\boldsymbol{\alpha}_1\varepsilon_{i-1}^2 + \boldsymbol{\alpha}_2)\ldots(\boldsymbol{\alpha}_1\varepsilon_1^2 + \boldsymbol{\alpha}_2|| < 0$, which results $||\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2|| < 1$ and $||(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)^j||$ summable. For some positive finite C_3,

$$\sum_{j=0}^{\infty}\sup(\sum_{k=0}^{j}\theta_1^k\theta_2^{j-k}) < \sum_{j=0}^{\infty}\sup(\theta_1+\theta_2)^j < \sum_{j=0}^{\infty}\sup(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)^j = C_3 < \infty.$$

Following (A.3), we further derive the autocovariance operator $\lambda_h(t,s)$ as,

$$\lambda_h = \mathbb{E}[(\sum_{j=0}^{\infty}(\sum_{k=0}^{j}\theta_1^k\theta_2^{j-k})\eta_{i-h-j})(\sum_{j=0}^{\infty}(\sum_{k=0}^{j}\theta_1^k\theta_2^{j-k})\eta_{i-j})]$$
$$= \sum_{j=0}^{\infty}\sum_{j=h}^{\infty}(\sum_{k=0}^{j}\theta_1^k\theta_2^{j-k})^2 \leq \sum_{j=0}^{\infty}\sum_{j=h}^{\infty}(\theta_1+\theta_2)^{2j} \leq C_4 < \infty.$$

Under Assumption A.2, the autocovariance converges to an upper bound C_4, which indicates an exponential decay for the autocovariance of an FARCH(2) process. Meanwhile, given the variance term $\lambda_0 = \sum_{j=0}^{\infty}\sum_{j=0}^{\infty}(\sum_{k=0}^{j}\theta_1^k\theta_2^{j-k})^2$, ρ_h^{farch} is derived. □

A.2. **Proof of Proposition 3.2.**

To prove Proposition 3.2, we assume an additional condition,

Assumption A.4. *There exists a bounded positive number κ, for $u = 1, \ldots, \kappa$, to let the kernels of $x_i(t)$ satisfying,*

$$\int \boldsymbol{\varphi}_j(t,s)\psi_u(s)ds \sim \mathfrak{p}_u(t)j^{-\varsigma}, \quad as \quad j \to \infty, \tag{A.4}$$

where the long-range parameter $1/2 < \varsigma < 1$, and the functions $\mathfrak{p}_u(t)$ satisfy the condition that their limit of $(\mathbb{E}\langle \mathfrak{p}_u v_i\rangle^2)^{1/2}$, for $u = 1, \ldots, \kappa$, is positively bounded and monotonically decreasing.

Assumption A.4 is a compact version of Assumption 2 stated in Li et al. (2019). Taking the form of the Karhunen-Loève representation, Assumption A.4 extends the limit results for scalar or multivariate long-range dependence into the functional time series context. This assumption is valid for most of empirical applications, and both the existence and consistency of the estimator of κ have been rigorously discussed in their paper. Also as indicated by Li et al. (2019), a long range dependent curve can be decomposed into a long-range dependence part and a short-range dependence part. Thus, Equation (A.4) shows that there exists a dominant sub-space, and projecting scores onto such a sub-space can reproduce the long-range dependence structure of $x_i(t)$, where the dependence degree is measured by the long-range dependence parameter ς.

Under the definition A.1, we know that, as a linear process, $x_i(t)$ is a weakly stationary functional time series with an MA(∞) representation. Decomposing $x_i(t)$ onto infinite dimensional spaces, there exist a sequence of orthonormal eigenfunction $\psi_u(t)$, $u \geq 1$ with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ that $x_i(t)$ can be approximated via,

$$x_i(t) = \sum_{u=1}^{\infty} \xi_i^u \psi_u(t),$$

where the functional score $\xi_i^u = \sum_{j=0}^{\infty} \int\int \boldsymbol{\varphi}_j(t,s) v_{i-j}(s) \psi_u(t) ds dt$, following (A.1).

To show the first part of Proposition 3.2, we have

$$\begin{aligned}
\mathbb{E}(\sigma_i^2) &= \mathbb{E}[\omega + \boldsymbol{\gamma}(x_{i-1}) + \boldsymbol{\alpha}(y_{i-1}^2) + \boldsymbol{\beta}(\sigma_{i-1}^2)] \\
&= \mathbb{E}[\omega + \boldsymbol{\gamma}(x_{i-1}) + (\boldsymbol{\alpha} + \boldsymbol{\beta})(y_{i-1}^2) - \boldsymbol{\beta}(\eta_{i-1})] \\
&= \mathbb{E}[\sum_{j=0}^{\infty} (\boldsymbol{\alpha} + \boldsymbol{\beta})^j (\omega + \boldsymbol{\gamma}(x_{i-j})) + \sum_{j=1}^{\infty} (\boldsymbol{\alpha} + \boldsymbol{\beta})^j \boldsymbol{\alpha}(\eta_{i-j})] \\
&= \sum_{j=0}^{\infty} (\boldsymbol{\alpha} + \boldsymbol{\beta})^j (\omega) + \sum_{j=0}^{\infty} (\boldsymbol{\alpha} + \boldsymbol{\beta})^j (x_{i-j}) < \infty,
\end{aligned}$$

where the covariate $x_i(t)$ is a stationary sequence under assumption A.4. Together with Assumption A.2 and the proof in A.1, the conditional volatility curve $\sigma_i^2(t)$ from the FGARCH-X model is well defined as $\sigma_i(t)^2 < \infty$, for all $i \geq 1$ and $t \in [0,1]$.

The second part of Proposition 3.2 can be shown straightforwardly. Using the definition of $x_i(t)$, we write the squared process of the FGARCH-X model in the following form,

$$\begin{aligned}
y_i^2 &= \sum_{j=0}^{\infty} \boldsymbol{\pi}^j (\omega + \boldsymbol{\gamma}(x_{i-j})) + \eta_i + \sum_{j=1}^{\infty} \boldsymbol{\pi}^{j-1} \boldsymbol{\alpha} \eta_{i-j}, \\
&= \sum_{j=0}^{\infty} \boldsymbol{\pi}^j (\omega) + \sum_{j=0}^{\infty} \boldsymbol{\pi}^j \boldsymbol{\gamma} (\sum_{k=0}^{\kappa} \boldsymbol{\varphi}_k (v_{i-j-k})) + \eta_i + \sum_{j=1}^{\infty} \boldsymbol{\pi}^{j-1} \boldsymbol{\alpha} \eta_{i-j} \qquad \text{(A.5)} \\
&= \sum_{j=0}^{\infty} \boldsymbol{\pi}^j (\omega) + \sum_{j=0}^{\infty} \sum_{k=0}^{\kappa} \boldsymbol{\pi}^j \boldsymbol{\gamma} \boldsymbol{\varphi}_k v_{i-j-k} + \eta_i + \sum_{j=1}^{\infty} \boldsymbol{\pi}^{j-1} \boldsymbol{\alpha} \eta_{i-j}.
\end{aligned}$$

This shows that y_i^2 is a linear process that is subject to iid variables v_i and η_i. By definition y_i^2 is short-range dependent if both $\sup || \sum_{k=0}^{\kappa} \boldsymbol{\pi}^j \boldsymbol{\gamma} \boldsymbol{\varphi}_k ||$ and $\sup ||\boldsymbol{\pi}^{j-1} \boldsymbol{\alpha}||$ are summable. From A.1 we have known that $\sum_{j=1}^{\infty} \sup ||\boldsymbol{\pi}^{j-1} \boldsymbol{\alpha}|| < \infty$, but under Assumption A.4, it is obvious that summability of the prior one is invalid.

Given the result in (A.5), we calculate the autocovariance operator as,

$$
\begin{aligned}
\lambda_h &= \mathbb{E}[(\sum_{j=0}^{\infty}\sum_{k=0}^{\kappa}\boldsymbol{\pi}^j\boldsymbol{\gamma}\boldsymbol{\varphi}_k v_{i-h-j-k} + \eta_i + \sum_{j=1}^{\infty}\boldsymbol{\pi}^{j-1}\boldsymbol{\alpha}\eta_{i-h-j}) \\
&\quad (\sum_{j=0}^{\infty}\sum_{k=0}^{\kappa}\boldsymbol{\pi}^j\boldsymbol{\gamma}\boldsymbol{\varphi}_k v_{i-j-k} + \eta_i + \sum_{j=1}^{\infty}\boldsymbol{\pi}^{j-1}\boldsymbol{\alpha}\eta_{i-j})] \\
&= \sum_{j=0}^{\infty}\sum_{j=h}^{\infty}\boldsymbol{\pi}^{2j}\boldsymbol{\gamma}^2\sum_{k=0}^{\kappa}\sum_{k=0}^{\kappa}\boldsymbol{\varphi}_k^2 + \sum_{j=1}^{\infty}\sum_{j=h}^{\infty}(\boldsymbol{\pi}^{j-1}\boldsymbol{\alpha})^2 \\
&\leq \sum_{j=0}^{\infty}\sum_{j=h}^{\infty}\boldsymbol{\pi}^{2j}\boldsymbol{\gamma}^2\sum_{k=0}^{\kappa}\sum_{k=0}^{\kappa}\boldsymbol{\varphi}_k^2 + C_2 \\
&\sim \sum_{j=0}^{\infty}\sum_{j=h}^{\infty}\boldsymbol{\pi}^{2j}\boldsymbol{\gamma}^2\sum_{k=0}^{\kappa}\sum_{k=0}^{\kappa}\mathfrak{p}_k(h)^{-2\varsigma} + C_2 \\
&\sim C_5 h^{-2\varsigma} + C_2 < \infty, \quad as \quad h \to \infty. \quad \square
\end{aligned}
$$

A.3. **FARCH** (∞) **representation of the FGARCH(p,q) model.**

This section derives an ARCH(∞) representation of the model (3.1). Under assumptions A.1-A.3, we denote $\underline{\sigma}_i^2 = \begin{bmatrix} \sigma_i^2(t) \\ \sigma_{i-1}^2(t) \\ \vdots \\ \sigma_{i-p+1}^2(t) \end{bmatrix}$, $\underline{y}_i^2 = \begin{bmatrix} y_i^2(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \dots & \boldsymbol{\beta}_{p+1} & \boldsymbol{\beta}_p \\ \boldsymbol{I}_{\mathcal{H}} & 0 & \dots & 0 & 0 \\ 0 & \boldsymbol{I}_{\mathcal{H}} & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & \boldsymbol{I}_{\mathcal{H}} & 0 \end{bmatrix}$, and the volatility equation in (3.1) can be rewritten as

$$
\underline{\sigma}_i^2 = \underline{\boldsymbol{\omega}} + \sum_{\ell=1}^{\infty}\mathfrak{a}_{i-\ell}\underline{y}_{i-\ell}^2 \tag{A.6}
$$

where $\underline{\boldsymbol{\omega}} = \sum_{l=0}^{\infty}\boldsymbol{B}^{\ell}\underline{\omega}(t)$, for $\underline{\omega}(t) = [\omega(t), 0, \dots, 0]^{\top}$, and the kernel operator $\mathfrak{a}_{i-\ell}$ gives

$$
\mathfrak{a}_{i-\ell} = \begin{cases} \underline{\boldsymbol{\alpha}}_{\ell}^{\top} + \sum_{j=1}^{\ell-1}\boldsymbol{B}^j\underline{\boldsymbol{\alpha}}_{\ell-j}^{\top}, 0 \leq i \leq q, \\ \sum_{j=1}^{q}\boldsymbol{B}^j\underline{\boldsymbol{\alpha}}_{\ell-j}^{\top}, i > q, \end{cases} \text{ for } \underline{\boldsymbol{\alpha}}_j = [\boldsymbol{\alpha_j}, 0, \dots, 0]^{\top}.
$$

To admit this representation, we vectorize the volatility equation in (3.1) as

$$\begin{bmatrix} \sigma_i^2(t) \\ \sigma_{i-1}^2(t) \\ \vdots \\ \sigma_{i-p+1}^2(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \dots & \boldsymbol{\beta}_{p+1} & \boldsymbol{\beta}_p \\ \boldsymbol{I} & 0 & \dots & 0 & 0 \\ 0 & \boldsymbol{I} & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & \boldsymbol{I} & 0 \end{bmatrix} \begin{bmatrix} \sigma_{i-1}^2(t) \\ \sigma_{i-2}^2(t) \\ \vdots \\ \sigma_{i-p}^2(t) \end{bmatrix} + \begin{bmatrix} \omega(t) + \sum_{j=1}^{q} \boldsymbol{\alpha}_j(y_{i-j}^2)(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

where $\boldsymbol{I}$ is an identity operator in $L^2[0,1]^2$. To present the above equation in a vector form, we write iteratively,

$$\begin{aligned} \underline{\sigma_i^2} &= \boldsymbol{\mathcal{B}}\underline{\sigma_{i-1}^2} + \underline{Y_{\omega_i}} \\ &= \boldsymbol{\mathcal{B}}(\boldsymbol{\mathcal{B}}\underline{\sigma_{i-2}^2} + \underline{Y_{\omega_{i-1}}}) + \underline{Y_{\omega_i}} \\ \cdots &= \sum_{\ell}^{\infty} \boldsymbol{\mathcal{B}}^{\ell} \underline{Y_{\omega_{i-\ell}}}. \end{aligned}$$

And by substituting $\underline{Y_{\omega_{i-\ell}}}$, we sequentially obtain

$$\underline{\sigma_i^2} = \sum_{l=0}^{\infty} \boldsymbol{\mathcal{B}}^{\ell}[\omega(t) + \sum_{j=1}^{q} \boldsymbol{\alpha}_j(y_{i-j}^2)(t), 0, \dots, 0]_{i-\ell}^{\top}. \tag{A.7}$$

By expanding Equation (A.7), an explicit representation is given as,

$$\begin{aligned} \underline{\sigma_i^2} = &\sum_{l=0}^{\infty} \boldsymbol{\mathcal{B}}^{l}\underline{\omega}(t) + \underline{\boldsymbol{\alpha}}_1^{\top}\underline{y}_{i-1}^2(t) + (\underline{\boldsymbol{\alpha}}_2 + \boldsymbol{\mathcal{B}}\underline{\boldsymbol{\alpha}}_1^{\top})\underline{y}_{i-2}^2(t) + \\ &\cdots + (\underline{\boldsymbol{\alpha}}_q + \sum_{j=1}^{q-1} \boldsymbol{\mathcal{B}}^{j}\underline{\boldsymbol{\alpha}}_{q-j}^{\top})\underline{y}_{i-q}^2(t) + \sum_{\ell=q+1}^{\infty}\sum_{j=1}^{q} \boldsymbol{\mathcal{B}}^{j}\underline{\boldsymbol{\alpha}}_{\ell-j}^{\top}\underline{y}_{i-\ell}^2(t) \end{aligned}$$

where $\underline{\omega} = [\omega(t), 0, \dots, 0]^{\top}$, $\underline{y_i^2} = [y_i^2(t), 0, \dots, 0]^{\top}$, and $\underline{\boldsymbol{\alpha}}_j = [\boldsymbol{\alpha}_j, 0, \dots, 0]^{\top}$. According to the above specification, we can see that the kernel coefficients for the ARCH parts are not entirely included until the lag q, and the kernel coefficients on the lags of $x_i^2(t)$ become exponentially varying after the lag q.

REFERENCES

Andersen, T. G., Bollerslev, T. (1997). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Journal of Finance 52*, 975–1005.

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

Barndorff-Nielsen, O. E., Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics*, *4*, 1–30.

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

Bollerslev, T., Patton, A. J., Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, *192*, 1–18.

Casas, I., Gao, J. (2008). Econometric estimation in long-range dependent volatility models: Theory and practice. *Journal of Econometrics 147*, 72–83.

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.

Charles, A., Darné, O. (2014). Volatility persistence in crude oil markets. *Energy policy*, *65*, 729–742.

Christensen, K., Podolskij, M. (2007). Realized range-based estimation of integrated variance. *Journal of Econometrics 141*, 323–349.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*, 174–196.

Didericksen, D., Kokoszka, P., Zhang, X. (2012). Empirical properties of forecasts with the functional autoregressive model. *Computational Statistics 27*, 285–298.

Ding, Z., Granger, C. W., Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance 1*, 83–106.

Engle, R. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics 17*, 425–446.

Fink, H., Fuest, A., Port, H. (2018). The impact of sovereign yield curve differentials on value-at-risk forecasts for foreign exchange rates. *Risks 6*, 84.

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

Fuertes, A. M., Izzeldin, M., Kalotychou, E. (2009). On forecasting daily stock volatility: The role of intraday information and market conditions. *International Journal of Forecasting*, *25*, 259–281.

Górecki, T., Hörmann, S., Horváth, L., Kokoszka, P. (2018). Testing normality of functional time series. *Journal of time series analysis 39*, 471–487.

Gorgi, P., Hansen, P. R., Janus, P., Koopman, S. J. (2019). Realized Wishart-GARCH: A score-driven multi-asset volatility model. *Journal of Financial Econometrics*, *17*, 1–32.

Han, H. (2015). Asymptotic properties of GARCH-X processes. *Journal of Financial Econometrics 13*, 188–221.

Hansen, P. R., Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?. *Journal of applied econometrics*, *20*, 873–889.

Hansen, P. R., Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics 24*, 127–161.

Hansen, P. R., Huang, Z., Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics 27*, 877–906.

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

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., Rice, G., Whipple, S. (2016). Adaptive bandwidth selection in the long run covariance estimator of functional time series. *Computational Statistics & Data Analysis 100*, 676–693.

Kargin, V., Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis 99*, 2508–2526.

Kearney, F., Shang, H. L. (2019). Uncovering predictability in the evolution of the WTI oil futures curve. *European Financial Management*, In preprint.

Kokoszka, P., Reimherr, M. (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis 34*, 116–129.

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.

Li, D., Robinson, P. M., Shang, H. L. (2019) Long-Range Dependent Curve Time Series. *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2019.1604362

Ma, F., Liao, Y., Zhang, Y., Cao, Y. (2019). Harnessing jump component for crude oil volatility forecasting in the presence of extreme shocks. *Journal of Empirical Finance*, *52*, 40–55.

McLeod, A. I. (1998). Hyperbolic decay time series. *Journal of Time Series Analysis 19*, 473–483.

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

Rice, G., Wirjanto, T., Zhao, Y. (2020a) Tests for conditional heteroscedasticity of functional data. *Journal of Time Series Analysis*, In press.

Rice, G., Wirjanto, T., Zhao, Y. (2020b) Forecasting Value at Risk via intra-day curves. *International Journal of Forecasting*, In press.

Sigg, C. D., Buhmann, J. M. (2008) Expectation-maximization for sparse and non-negative pca. *Proceedings of the 25th International Conference on Machine Learning*. *ICML'08*, 960–967. New York, NY, USA, 2008. ACM.

Spangenberg, F. (2013) Strictly stationary solutions of ARMA equations in Banach spaces. *Journal of Multivariate Analysis 121*, 127–138.

Sun, H., Yu, B. (2020). Volatility asymmetry in functional threshold GARCH model. *Journal of Time Series Analysis 41*, 95–109.

Wei, Y., Wang, Y., Huang, D. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, *32*, 1477–1484.

Zhang, Y. J., Wang, J. L. (2019). Do high-frequency stock market data help forecast crude oil prices? Evidence from the MIDAS models. *Energy Economics*, *78*, 192–201.