

FORECASTING VALUE AT RISK VIA INTRA-DAY RETURN CURVES

GREGORY RICE¹, TONY WIRJANTO¹ AND YUQIAN ZHAO²

¹*Department of Statistics and Actuarial Science, University of Waterloo, Canada, N2L 3G1*

²*Essex Business School, University of Essex, UK, CO4 3SQ*

ABSTRACT. Methods for incorporating high resolution intra-day asset price data into risk forecasts are being developed at an increasing pace. Existing methods, such as those based on realized volatility, rely primarily on reducing the observed intra-day price fluctuations to simple scalar summaries. In this paper, we propose several methods that incorporate full intra-day price information as functional data objects in order to forecast Value at Risk (VaR). The methods we consider are based on the recently proposed functional GARCH models and a new functional linear quantile regression model. In addition to providing daily VaR forecasts, these methods can also be used to forecast intra-day VaR curves, which we develop and study along with companion backtests to evaluate the quality of such intra-day risk measures. Using high-frequency trading data from equity and foreign exchange markets, we first forecast 1-day-ahead daily and intra-day VaR by applying the proposed methods and a host of benchmark models. Empirical evidence suggests that the functional GARCH models estimated from over-night cumulative intra-day return curves exhibit competitive performance relative to benchmark models in daily risk management, and produce valid intra-day VaR curves.

JEL Classification: C12, C32, C58, G10

Keywords: Value at Risk, Overnight Cumulative Intra-day Return, Functional GARCH, Intra-day VaR Backtesting, Forecasting Comparison.

1. INTRODUCTION

One of the primary concerns for large financial institutions and regulators is to forecast risk measures, such as Value at Risk (VaR), accurately and efficiently. These specific measures have

E-mail address: Corresponding author: y.zhao@essex.ac.uk.

been brought to a prominent status as a result of the Basel II Accord in 1996. Suppose that Y_i denotes the daily return of an asset, and further suppose for the sake of simplicity that the map $x \rightarrow \mathbb{P}(Y_i \leq x | \mathcal{F}_{i-1})$ is strictly increasing with probability 1, where \mathcal{F}_{i-1} denotes the previous information set. The VaR at quantile τ , VaR_i^τ , is defined to be the almost surely unique \mathcal{F}_{i-1} -measurable random variable satisfying

$$(1.1) \quad \mathbb{P}(Y_i \leq \text{VaR}_i^\tau | \mathcal{F}_{i-1}) = \tau,$$

so that VaR_i^τ quantifies a potential market loss of an asset on day i that is not expected to be exceeded over a certain period with probability τ . Assuming strictly positive prices, one can equivalently model the VaR where Y_i in (1.1) is the log returns by performing an appropriate transformation, see Example 12.6 of Francq and Zakoian (2011); this is how we proceed in this paper. Forecasting VaR for daily mark-to-market portfolios is routinely used to facilitate an internal risk control from the bank side in an attempt to contribute to a stable environment for financial markets.

Quantifying the conditional volatility of Y_i is a critical step in VaR forecasting, and this will evidently depend on the available information contained in \mathcal{F}_{i-1} . The vast majority of existing studies on VaR forecasting assumes that information is only available on the returns at a daily frequency; see Nieto and Ruiz (2016) for a recent survey on the subject. However, researchers often have access to much more refined price data on the asset of interest, such as high-frequency intra-day records. There is every reason to believe that this additional information would be relevant for predicting the conditional volatility of Y_i by supplementing the daily information. For example, given that an asset return on a given day was zero, one might expect greater volatility on the following day if the intra-day price had high fluctuations compared to the case if there were only limited intra-day fluctuations.

The problem that we wish to study in this paper is how to incorporate this intra-day information into daily VaR forecasts. This problem has begun to receive attention only relatively recently, and the methods proposed in the literature to date rely primarily on using realized measures of the intra-day volatility in the forecast. Such realized measures are, roughly speaking, scalar summaries of the intra-day price fluctuations; see, e.g. Andersen et al. (2003).

Based on realized measures such as the quadratic variation and the intra-day high minus the intra-day low, early work in this stream of research showed that incorporating this information only provides a small improvement on VaR forecasts over simple generalized autoregressive conditionally heteroscedastic (GARCH) model based benchmarks (Giot and Laurent, 2004; Angelidisa and Degiannakis, 2008). Angelidisa and Degiannakis (2008) attributed this to the distinct and apparently incompatible underlying processes driving intra-day and daily returns. Subsequently, realized GARCH (RGARCH) models have been introduced to allow the dynamics of daily conditional volatility to be determined by past intra-day realized volatility (Hansen et al., 2012), while as a GARCH-type model, it also imposes distributional restrictions to describe the dynamics of returns. Seeking a parsimonious model for incorporating volatility dynamics and different temporal frequencies, Corsi (2009) proposed a Heterogeneous AutoRegressive model of Realized Volatility (HAR-RV), which has been widely used to forecast volatility. More recently, a number of empirical papers (e.g., Gerlach and Wang, 2016; Meng and Taylor, 2018) argue convincingly that these models significantly improve VaR forecasts over benchmark models. Although the realized volatility-type models exploit a connection between intra-day and daily frequency returns, they still only use a scalar summary of the evolution of the intra-day price that leads to a given return.

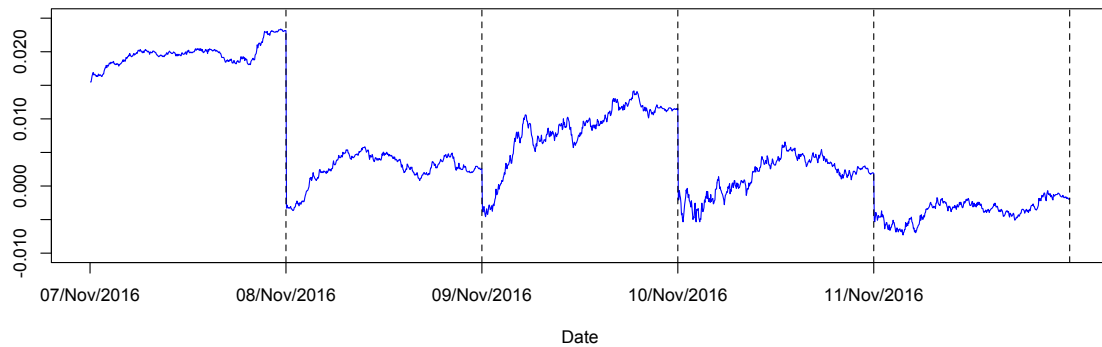
An alternative way of incorporating all of the intra-day price data into a VaR forecast is to treat the high-frequency intra-day asset price records as discrete observations from an underlying daily price or return *curve*. Such curves can be constructed to contain information on the daily returns, intraday price movements, over-night returns, etc., and may subsequently be modeled or used as covariates in order to quantify conditional heteroscedasticity, or estimate the conditional quantiles, of the returns. We coin the particular curve used to accomplish this task an *over-night cumulative intra-day return* (OCIDR) curve, $X_i(t)$, defined as

$$(1.2) \quad X_i(t) = \log P_i(t) - \log P_{i-1}(1), \quad t \in [0, 1], \quad 1 \leq i \leq N$$

where $P_i(t)$ is the price on day i at intra-day time t , with t normalized to the interval $[0, 1]$. With this normalization $P_i(1)$ represents the closing price on day i , and so the sequence $\{X_i(1), i \in \mathbb{Z}\}$ is the daily Y_i . A set of similar curves termed *cumulative intra-day return* (CIDR) curves were introduced in Gabrys et al. (2010) and their properties have been studied

extensively in a number of papers, which include Kokoszka and Zhang (2012) and Kokoszka and Reimherr (2013). We instead prefer to use OCIDR curves because when $t = 0$ equation (1.2) evaluates the over-night return $\log P_i(0) - \log P_{i-1}(1)$, which has been shown to be an important element in forecasting volatility (Hansen and Lunde, 2006; Meng and Taylor, 2018). Although these curves are in practice only observed at an arbitrarily fine grid of intraday times, these may be easily interpolated to produce full curves. As an example, Figure 1.1 exhibits a plot of the OCIDR curves of the Standard & Poor's (S&P)500 index during the week of the 58th U.S. presidential election in 2016, where the curves are generated using linear interpolation.

FIGURE 1.1. Plot of five OCIDR Curves from November 2016 derived from one-minute resolution records of the S&P500 index.



In order to incorporate the curves $X_i(t)$ in VaR forecasts, we employ new methods in functional time series analysis (FTSA). For book length treatments of this subject, see, for instance, Bosq (2000), Ramsay and Silverman (2006), Ferraty and Vieu (2006), and Horváth and Kokoszka (2012). This work builds upon several other recent papers that use tools from FTSA in order to forecast risk measures. For instance, Ferraty and Quintela-del-Río (2016) proposed a non-parametric kernel-based FTSA method to estimate the conditional VaR and Expected Shortfall (ES) based on daily return data. Fink et al. (2018) considered forecasting VaR by employing scalar ARMA-GARCH models with functional exogenous covariates in foreign exchange markets; see also Fuest et al. (2015). Horta and Ziegelmann (2018) proposed a functional time series approach to forecast the probability density of intraday returns towards estimating daily VaR.

The main method considered in this paper entails modeling the entire sequence of curves $X_i(t)$ using functional GARCH type models, which yield as a byproduct forecasts of the conditional

quantile of the log returns $X_i(1)$. This method makes use of a number of recent advances on modeling functional conditional heteroscedasticity. A functional generalization of Engle's (1982) original ARCH model was put forward in Hörmann et al. (2013), and has subsequently been generalized to functional GARCH models in Aue et al. (2016) and Cerovecki et al. (2019). Goodness of fit tests for these models that may be used to aid in their specification were proposed in Rice et al. (2019). Having estimated such a model, VaR forecasts may be obtained by specifying or estimating a distribution for the residuals, and for this we consider several methods that include functional extreme value processes and bootstraps.

In addition, the FGARCH models yield quantile curves at a continuous time scale, which we term VaR curves, from which daily VaR forecasts are produced as a simple byproduct. To give a brief description, we may use such models to forecast the pointwise (i.e., for a fixed quantile level) conditional τ th quantile $\text{VaR}_i^\tau(t)$ of $X_i(t)$, for $t \in [0, 1]$. The standard daily VaR is then obtained as $\text{VaR}_i^\tau(1)$, while $\text{VaR}_i^\tau(t)$ describes how the risk is forecast to evolve over the course of the trading day, and can be used as a tool to quantify intra-day risk. Below, we illustrate this application, and also develop novel backtests inspired by Christoffersen (1998) to evaluate the quality of these intra-day VaR curves. Such tests may also be useful for assessing other functional data-based risk measures.

In order to provide a more complete study of using functional time series techniques to forecast daily VaR, we also propose to estimate the conditional quantile of the returns via a new functional linear quantile regression model in which the functions $X_i(t)$ act as covariates in order to predict the quantile of returns. However, such a model shows inferior empirical performance, we thus relegate this into the Online supplement.

Each of the proposed models were evaluated by forecasting and backtesting the one-day-ahead VaR for three equity indices, S&P500, DAX30 and CAC40, and one foreign exchange market (FOREX) example, USD-Euro, ranging from January 2012 to March 2018. These are further compared to a number of benchmark models, such as GARCH filtered historical simulation (FHS) models, RGARCH models, and HAR-RV models with normally distributed or bootstrapped innovations. The results showed that, although the benchmark models with bootstrapped innovations provide good VaR forecasts, the FGARCH-type methods incorporating the OCIDR curves exhibited competitive and in some cases better forecasting performance,

while also producing valid forecasts of the VaR at an intra-day scale relevant to real-time risk management.

The rest of the paper is structured as follows. Section 2 introduces the functional GARCH model. Section 3 discusses the results of the empirical analyses, and Section 4 concludes the paper.

Throughout the paper, we use the following notation. We let $L^2[0, 1]^d$ to denote the space of functions defined on the unit hypercube in d dimensions with a finite squared integral. The random norm of X is denoted and defined as $\|X\| = [\int_0^1 X^2(t)dt]^{1/2}$, based on the inner product $\langle X_1, X_2 \rangle = \int_0^1 X_1(t)X_2(t)dt$. We write \int_0^1 as \int for convenience.

2. VaR FORECASTING USING FUNCTIONAL GARCH MODELS

As specified in (1.2), we assume that we have observed N OCIDR curves $X_1(t), \dots, X_N(t)$, and we are interested in forecasting the VaR, or the conditional quantile, at level τ of the log returns $Y_{N+1} = X_{N+1}(1)$, as well as the pointwise conditional τ th quantile of $X_{N+1}(t)$. Practically speaking, on a given day i one does not observe the full price curve $P_i(t)$ needed to define $X_i(t)$, instead $P_i(t)$ is observed usually at a fine grid of intraday times points; for instance one observation every minute. We discuss below how to interpolate these points in order to produce full price curves, but for now we assume that the OCIDR curves are fully observed.

The first method that we propose in this paper is a two step procedure in which first the entire series $X_i(t)$ is modeled using a functional GARCH model, after which the distribution of the residuals is either estimated or bootstrapped in order to produce the forecasts of interest. Following Aue et al. (2017) and Cerovecki et al. (2019), we assume that $X_i(t)$ has the following specification:

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

where $\mu(t)$ is the mean function of the sequence $X_i(t)$, the innovation functions $\varepsilon_i(t)$ form an independent and identically distributed sequence, and $\sigma_i(t)$ defines the intra-day conditional standard deviation curve. We let σ_i^2 be measurable with respect to \mathcal{F}_{i-1} , where in this discussion of FGARCH models we assume that \mathcal{F}_i is the sigma-algebra generated by ε_j , $j \leq i$, and this will be incorporated in certain modeling specifications of σ_i^2 that we detail below. In practice, it

is either assumed that the mean curve $\mu(t)$ is zero, or demeaned curves are used in the analysis, and we use the latter of these two choices in this paper. Following Equation (1.1), we can introduce the intra-day τ th VaR curve VaR_i^τ through the random variables $\text{VaR}_i^\tau(t)$, $t \in [0, 1]$, defined as the almost surely unique \mathcal{F}_{i-1} -measurable solution to random variable satisfying

$$(2.2) \quad \mathbb{P}(X_i(t) \leq \text{VaR}_i^\tau(t) | \mathcal{F}_{i-1}) = \tau,$$

where again for convenience of notation we assume that the conditional distribution of $X_i(t)$ on \mathcal{F}_{i-1} is strictly increasing. Under model (2.1), the pointwise 1-day-ahead τ th conditional quantile of the curve $X_i(t)$ is given by

$$(2.3) \quad \text{VaR}_i^\tau(t) = \sigma_i(t) \varepsilon^\tau(t),$$

where $\varepsilon^\tau(t)$ is the τ th quantile of $\varepsilon_1(t)$ conditional on \mathcal{F}_0 , given the result that the conditional quantiles of ε are time-invariant under the first-order or strict stationarity. We note that the intra-day $\text{VaR}_i^\tau(t)$ in (2.3) is an intra-day generalization of VaR_i^τ , and $\text{VaR}_i^\tau = \text{VaR}_i^\tau(1)$. A forecast for the VaR curve $\text{VaR}_{N+1}^\tau(t)$ can then be obtained as

$$(2.4) \quad \widehat{\text{VaR}}_{N+1}^\tau(t) = \widehat{\sigma}_{N+1}(t) \widehat{\varepsilon}^\tau(t),$$

where $\widehat{\sigma}_{N+1}(t)$ is a forecast of $\sigma_{N+1}(t)$, and $\widehat{\varepsilon}^\tau(t)$ is an estimate of the τ th quantile of the error process.

In order to forecast $\sigma_{N+1}(t)$, we assume that it follows an FGARCH(p, q) specification of the form

$$\sigma_i^2(t) = \omega(t) + \sum_{k=1}^q \alpha_k X_{i-k}^2(t) + \sum_{j=1}^p \beta_j \sigma_{i-j}^2(t),$$

restricting our attention to the FGARCH(1,1) model, the model reduces to,

$$(2.5) \quad \sigma_i^2(t) = \omega(t) + \alpha_1 X_{i-1}^2(t) + \beta_1 \sigma_{i-1}^2(t),$$

where the operators α_1 and β_1 are kernel integral operators satisfying $\alpha_1 X_{i-1}^2(t) = \int \alpha_1(t, s) X_{i-1}^2(s) ds$ and $\beta_1 \sigma_{i-1}^2(t) = \int \beta_1(t, s) \sigma_{i-1}^2(s) ds$, $t \in [0, 1]$. We assume a positive function $\omega \in L^2[0, 1]$,

and non-negative kernel functions $\alpha_1(t, s)$ and $\beta_1(t, s)$ defining the operators are assumed to be elements of $L^2[0, 1]^2$.

Cerovecki et al. (2019) give precise conditions in terms of the innovation sequence, ω , and the α_1 's and β_1 's under which a stationary solution to (2.1) and (2.5) exists, and we assume that these conditions hold in order for the estimation procedures that we define below to be consistent. The notation required to reproduce these conditions here is prohibitively cumbersome, but roughly speaking they require that the functions ω , α_1 's, and β_1 's are non-negative in order to guarantee the positivity of $\sigma_i^2(t)$, and that the operators α_1 and β_1 satisfy contraction conditions that are typically made in multivariate GARCH models.

Under such conditions the model may be estimated by projecting the intercept $\omega(t)$ into a finite, M_1 dimensional space, as well as the kernels $\alpha_1(t, s)$ and $\beta_1(t, s)$ into a subspace of $L^2[0, 1]^2$ of dimension M_1^2 . Subsequently using least squares estimation (LSE) or quasi-maximum likelihood estimation (QMLE) the parameters in the dimension-reduced model can be consistently estimated; see Aue et al. (2017) and Cerovecki et al. (2019), respectively. Consider an orthonormal collection of basis functions $\hat{\phi}_1(t), \dots, \hat{\phi}_{M_1}(t)$ in $L^2[0, 1]$ that we use to perform the dimensionality reduction, where we use “hat” notation to indicate that these functions might depend on the data. We will discuss the selection of these elements in detail momentarily. The parameter functions defined in (2.5) are then assumed to be approximated via a fixed set of basis functions,

$$\omega(t) \approx \sum_{i=1}^{M_1} d_i \hat{\phi}_i(t), \quad \alpha_1(t, s) \approx \sum_{i,j=1}^{M_1} a_{i,j} \hat{\phi}_i(t) \hat{\phi}_j(s), \quad \beta_1(t, s) \approx \sum_{i,j=1}^{M_1} b_{i,j} \hat{\phi}_i(t) \hat{\phi}_j(s),$$

where d_i are elements in the positive valued vector $D \in \mathbb{R}^{M_1}$, $a_{i,j}$ and $b_{i,j}$ are the (i, j) th entries in non-negative valued matrices A_1 and B_1 in $\mathbb{R}^{M_1 \times M_1}$, respectively.

Given the basis elements and up to this approximation, the model is fully determined by the parameter $\theta = \text{vec}(D, A_1, B_1)$ which is assumed to lie in a compact subset $\Theta \subset \mathbb{R}^{M_1 + 2M_1^2}$. Let the projecting vectors $\mathbf{v}_i^2 = [\langle X_i^2, \hat{\phi}_1 \rangle, \dots, \langle X_i^2, \hat{\phi}_{M_1} \rangle]^\top$, and $\mathbf{s}_i^2 = [\langle \sigma_i^2, \hat{\phi}_1 \rangle, \dots, \langle \sigma_i^2, \hat{\phi}_{M_1} \rangle]^\top$. By setting $\mathbf{s}_1^2 = [d_1, \dots, d_{M_1}]^\top$, we can recursively compute $\mathbf{s}_i^2(\theta)$ as,

$$\mathbf{s}_i^2(\theta) = D + A_1 \mathbf{v}_{i-1}^2 + B_1 \mathbf{s}_{i-1}^2.$$

Iterating the above equation yields,

$$\mathbf{s}_i^2(\theta) = \sum_{l=1}^{\infty} B^{l-1} D + \sum_{l=1}^{\infty} B^{l-1} A_1 \mathbf{v}_{i-1}^2.$$

Then the parameter vector may be consistently estimated via the LSE approach as

$$(2.6) \quad \hat{\theta}_N = \underset{\theta' \in \Theta}{\operatorname{argmin}} \left\{ \sum_{i=2}^N (\mathbf{v}_i^2 - \hat{\mathbf{s}}_i^2(\theta'))^\top (\mathbf{v}_i^2 - \hat{\mathbf{s}}_i^2(\theta')) \right\},$$

This estimator in turn gives estimates of the functions,

$$\hat{\omega}(t) = \sum_{i=1}^{M_1} \hat{d}_i \hat{\phi}_i(t), \quad \hat{\alpha}_1(t, s) = \sum_{i,j=1}^{M_1} \hat{a}_{i,j} \hat{\phi}_i(t) \hat{\phi}_j(s), \text{ and } \hat{\beta}_1(t, s) = \sum_{i,j=1}^{M_1} \hat{b}_{i,j} \hat{\phi}_i(t) \hat{\phi}_j(s),$$

and hence also estimators $\hat{\alpha}_1$ and $\hat{\beta}_1$ of the operators α_1 and β_1 . The estimated conditional variance curves are then given by

$$(2.7) \quad \hat{\sigma}_{i+1}^2(t) = \hat{\omega}(t) + \hat{\alpha}_1 X_i^2(t) + \hat{\beta}_1 \hat{\sigma}_i^2(t),$$

with $\hat{\sigma}_{N+1}^2(t)$ giving the one-step ahead forecast.

The basis $\hat{\phi}_1(t), \dots, \hat{\phi}_{M_1}(t)$ can be selected in a number of ways, including using deterministic bases, such as the fourier bases, b-spline bases or wavelet bases. Here we utilize a functional principal component analysis (FPCA) in order to obtain data-driven basis functions; see Ramsay and Silverman (2006) for a detailed monograph on the FPCA technique. As $\alpha_1(t, s)$ and $\beta_1(t, s)$ are the kernel functions relating to the conditional mean of the squared process $X_i^2(t)$, we consider the basis of eigenfunctions of the covariance operator of $X_i^2(t)$ in order to perform a dimensionality reduction. The covariance kernel of the squared process can be estimated from the sample as

$$\hat{c}(t, s) = \frac{1}{N} \sum_{i=1}^N [X_i^2(t) - \bar{X}_i^2(t)][X_i^2(s) - \bar{X}_i^2(s)],$$

where $\bar{X}_i^2(t) = \frac{1}{N} \sum_{i=1}^N X_i^2(t)$. Under the simple moment condition that $E\|X_i\|^4 < \infty$, the above kernel is a consistent estimator of $c(t, s) = \operatorname{cov}(X_i^2(t), X_i^2(s))$. The principal components of the squared process can then be consistently estimated by $\hat{\phi}_i$ which solve the Fredholm equations

$$\hat{\lambda}_i \hat{\phi}_i(t) = \int \hat{c}(t, s) \hat{\phi}_i(s) ds, \quad \forall i.$$

Meanwhile, because these bases are used to model volatility, the conditions of positive $\widehat{\omega}(t)$ and non-negative $\widehat{\alpha}_1(t, s), \widehat{\beta}_1(t, s)$ should be satisfied to guarantee the positivity of $\widehat{\sigma}_i^2(t)$. In order to satisfy this condition, inspired by the discussion on the positivity in Cerovecki et al. (2019), we set the negative part of the estimated bases to zero, i.e. we take

$$\widehat{\phi}_i(t) = \widehat{\phi}_i(t) - \inf_{t \in [0,1]} \widehat{\phi}_i(t) \wedge 0, \quad \forall i.$$

As a consequence, the constant function $\omega(t)$ in Equation (2.5) is positive as the vector parameters D are positive valued, and the kernel functions $\alpha_1(t, s)$ and $\beta_1(t, s)$ are non-negative as the parameter matrices A_1 and B_1 are non-negative. Below we present results based on this basis selection. We considered several other polynomial bases, e.g., B-spline bases, but found this particular choice to give superior results.

A practical question that also must be addressed is how to select p, q and M_1 to fit a given collection of curves. Various methods have been suggested in order to choose the projection number M_1 . For instance, in the usual vernacular of principal component analysis, we can select M_1 as the minimal number of basis elements so that the total variance explained (TVE) by them is a high percentage of that of the squared processes, e.g. 90%. In this paper, we propose to use a global evaluation of model fit in order to select p, q and M_1 simultaneously, following the work of Rice et al. (2019). Let the residuals be defined as $\widehat{\varepsilon}_i(t) = X_i(t)/\widehat{\sigma}_i(t)$ for a given selection of p, q , and M_1 . We argue that the model is adequate if we cannot reject the null hypothesis

\mathcal{H}_0 : The sequence $\{\varepsilon_i(t)\}$ is independent, and identically distributed (IID).

based on the residuals. To test \mathcal{H}_0 we apply the following two statistics that aim to detect conditional heteroscedasticity of the residuals:

$$(2.8) \quad V_{\widehat{\varepsilon}, N, H} = N \sum_{h=1}^H \widehat{\gamma}_{1, \widehat{\varepsilon}, h}^2, \quad Q_{\widehat{\varepsilon}, N, H} = N \sum_{h=1}^H \|\widehat{\gamma}_{2, \widehat{\varepsilon}, h}\|^2,$$

where H is a user-specified maximal lag length,

$$\widehat{\gamma}_{1, \widehat{\varepsilon}, h} = \frac{1}{N} \sum_{i=1}^{N-h} (\|\widehat{\varepsilon}_i\|^2 - \bar{\varepsilon}_1)(\|\widehat{\varepsilon}_{i+h}\|^2 - \bar{\varepsilon}_1),$$

and

$$\widehat{\gamma}_{2,\widehat{\varepsilon},h}(t,s) = \frac{1}{N} \sum_{i=1}^{N-h} (\widehat{\varepsilon}_i^2(t) - \bar{\varepsilon}_2(t))(\widehat{\varepsilon}_{i+h}^2(s) - \bar{\varepsilon}_2(s)),$$

where $\bar{\varepsilon}_1$ and $\bar{\varepsilon}_2(t)$ are sample means of $\|\widehat{\varepsilon}_i\|^2 \in \mathbb{R}$ and $\widehat{\varepsilon}_i^2 \in L^2[0,1]$, respectively. The asymptotic distributions of each of these statistics are calculated in Rice et al. (2019), which can be used to calculate an approximate P-value for \mathcal{H}_0 , with small P-values indicating inadequacy of the model. The parameters p , q and M_1 can be increased as needed until a model is found to be adequate.

2.1. Estimating $\varepsilon^\tau(t)$.

Once the conditional volatility forecast $\widehat{\sigma}_{N+1}(t)$ has been obtained, one only has to specify the τ th quantile of intra-day innovations $\varepsilon_i^\tau(t)$ in order to produce an estimate for the VaR. Here we enumerate three methods to do so. The first we denote with “FGARCH-OU”, in which we assume that the innovation curves $\varepsilon_i(t)$ follow an Ornstein-Uhlenbeck process,

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

where $W_i(s)$ are independent and identically distributed standard Brownian motions. It follows from this that $\varepsilon_i(t)$ for each t are distributed as standard normal random variables, and their τ th quantile is given by $\widehat{\varepsilon}^\tau(t) = \Phi^{-1}(\tau)$, where $\Phi(\cdot)$ is the standard normal distribution function. In comparison to standard univariate GARCH-based VaR forecasts, this is similar to assuming an independent and identically distributed standard normal innovation sequence.

Secondly, as the assumption of normality is usually in question with financial return data, we propose to resample the residuals $\widehat{\varepsilon}_i(t)$ to empirically estimate the quantiles $\widehat{\varepsilon}^\tau(t)$ pointwise. We denote this method “FGARCH-BOOT”, and it is implemented as follows. For j ranging from 1 to B , B denoting the number of bootstrap samples, and letting $\varepsilon_i^{(j)}(t)$ denote the j 'th draw with replacement from the residual curves, we calculate the j 'th bootstrap realization of $X_{N+1}(t)$ as $\widehat{X}_{N+1}^{(j)}(t) = \widehat{\sigma}_{N+1}(t) \varepsilon_i^{(j)}(t)$. The VaR may then be estimated as $\widehat{\text{VaR}}_{N+1}^\tau(t) = \widehat{X}_{N+1}^{S,\tau}(t)$, where $\widehat{X}_{N+1}^{S,\tau}(t)$ is the pointwise sample τ th bootstrap quantile. For one step/day ahead forecast, this is approximately equivalent to taking

$$(2.9) \quad \widehat{\varepsilon}^\tau(t) = \widehat{\varepsilon}^{S,\tau}(t),$$

in (2.4), where $\widehat{\varepsilon}^{S,\tau}(t)$ is the point-wise sample τ th quantile of $\widehat{\varepsilon}_i(t)$. For a multiple-day ahead VaR forecast one can resample the residuals further to produce bootstrap realizations of $X_{N+h}(t)$, and estimate the VaR similarly.

Thirdly, in order to more accurately describe the tail of the distribution of the innovations, we consider models based on extreme value theory (EVT), which we denote as “FGARCH-EVT”. This method is inspired by McNeil and Frey (2000), in which the VaR is forecast by estimating an extreme value distribution for scalar GARCH innovations. In their work they found a generalized Pareto distribution (GPD) to be suitable. To adapt this idea to our context, we first project the residuals $\widehat{\varepsilon}_i(t)$ onto a finite, M_2 -dimensional subspace spanned by an orthonormal set $\widehat{\phi}_{\varepsilon,m}$, $1 \leq m \leq M_2$, where the $\widehat{\phi}_{\varepsilon,m}(t)$ are eigenfunctions of the empirical covariance operator of the residuals $\widehat{\varepsilon}_i$. Then, we approximate

$$(2.10) \quad \widehat{\varepsilon}_i(t) \approx \sum_{m=1}^{M_2} \zeta_{m,i} \widehat{\phi}_{\varepsilon,m}(t).$$

For each fixed m , we estimate an extreme value distribution from the sample $\zeta_{m,1}, \dots, \zeta_{m,M_2}$. A procedure similar to this was considered recently in Kokoszka and Xiong (2018). We consider the peaks over threshold (POT) model (Gomes and Guillou, 2015) with threshold u_m , to be discussed below, for univariate $\zeta_{m,i}$, and fit exceedances with the GPD.

The value of a threshold u_m is critical to the POT model. As discussed in Gomes and Guillou (2015) there is a well-known trade-off between selecting a threshold close to the center of the distribution or in the tail. The former can result in an estimate of GPD that fails to mimic the tail behavior, while the latter can result in a low effective sample size as very few observations exceed the threshold. Some studies have applied graphical methods to select u_m , but ultimately the choice is still subjective. Here, we stick with the approach taken by McNeil and Frey (2000) and let $P(\zeta_{m,i} < u_m) = 0.05$, i.e., the threshold u_m is determined to keep the number of exceedances to be always equal to 5% of the total number of observations. With this specification, we then fit each sequence of scores $\zeta_{m,i}$ with a GPD, which has the following density function,

$$(2.11) \quad G_{\gamma_m, \nu_m}(\zeta) = \begin{cases} 1 - (1 + \gamma_m \zeta / \nu_m)^{-1/\nu_m} & \text{if } \gamma_m \neq 0 \\ 1 - \exp(-\zeta / \nu_m) & \text{if } \gamma_m = 0 \end{cases},$$

where the scale parameter ν_m and the shape parameter γ_m can be estimated by using the maximum likelihood method. Using the estimated parameters, we are able to simulate random scores $\zeta_{m,i}^{(j)}$, $j = 1, \dots, B$, yielding a sequence of simulated innovations

$$\widehat{\varepsilon}_i^{(j)}(t) = \sum_{m=1}^{M_2} \zeta_{m,i}^{(j)} \widehat{\phi}_m(t).$$

based on a large number of such simulations, we obtain a Monte Carlo pointwise τ th quantile function $\widehat{\varepsilon}^\tau(t)$ to use in (2.4).

2.2. Estimating and Backtesting Intra-day VaR Curves.

As mentioned above, the FGARCH model can be used to forecast the standard daily VaR with the forecast $\widehat{\text{VaR}}_{N+1}^\tau(1)$ via Equation (2.4). In addition, the function $\widehat{\text{VaR}}_{N+1}^\tau(t)$, $t \in [0, 1]$, may be used to forecast the intra-day Value at Risk, and provides a visual and implementable tool for intra-day risk management.

In order to evaluate the quality of such functional VaR forecasts, we propose a statistical approach to backtest the curves $\widehat{\text{VaR}}_{N+1}^\tau(t)$. Inspired by classical backtesting approaches (Christoffersen, 1998; 2010; 2011), we base these backtests on the pointwise intra-day violation process $Z_i^\tau(t)$ at quantile τ , defined by

$$(2.12) \quad Z_i^\tau(t) = \mathbb{1}_{\{X_i(t) < \widehat{\text{VaR}}_i^\tau(t)\}}, \quad 1 \leq i \leq N, \quad t \in [0, 1],$$

where $X_i(t)$ is the observed OCIDR return curve, and $\mathbb{1}$ is the indicator function. In conventional VaR backtesting methods, as surveyed in Christoffersen (2010), one usually wishes to evaluate (1) that the conditional coverage is at the specified nominal level τ , and (2) that the violation process is approximately independent. This evaluation can be framed as hypotheses tests of

$$\mathcal{H}_0^{(1)}: \mathbb{E}[Z_i^\tau(t) - \tau] = 0 \text{ for all } t \in [0, 1], \text{ and } \quad \mathcal{H}_0^{(2)}: Z_i^\tau(t) \text{ is IID along } i.$$

Unlike Christoffersen (1998), who used likelihood ratio-based statistics to jointly test such hypotheses, here we propose tests based on some recently developed techniques in functional time series analysis.

In particular, in order to test $\mathcal{H}_0^{(1)}$, we use the test statistic

$$T_N = N \|\bar{Z} - \tau\|^2,$$

where $\bar{Z}(t) = 1/N \sum_{i=1}^N Z_i(t)$. Under $\mathcal{H}_0^{(2)}$ it is straightforward to verify that

$$(2.13) \quad T_N \xrightarrow{\mathcal{D}} \sum_{\ell=1}^{\infty} \lambda_{\ell} \mathcal{N}_{\ell}^2 \stackrel{\mathcal{D}}{=} \Xi_1, \quad N \rightarrow \infty,$$

where $\mathcal{N}_{\ell}, \ell \geq 1$ are independent and identically distributed normal random variables, and the $\lambda_{\ell}, \ell \geq 1$ are the eigenvalues of the covariance operator of the violation process, namely they are the eigenvalues of the kernel integral operator with kernel $C_Z(t, s) = \text{cov}(Z_0(t), Z_0(s))$, $t, s \in [0, 1]$. The distribution of Ξ_1 can be approximated by using Monte Carlo simulation after estimating the eigenvalues λ_{ℓ} , and a P-value of the test may then be approximated by $p = P(\Xi_1 > T_N)$.

We note here that it is a typical simplification in the literature on conditional coverage backtests to assume that the violation process follows $\mathcal{H}_0^{(2)}$ or a martingale difference assumption in order to establish and use results akin to (2.13), although similar results evidently hold under weaker conditions. In a preliminary study, we also considered backtests for nominal conditional coverage that allowed for serial correlation in the violation process, but found that these had little effect on the empirical results presented below, and so we chose not to develop them here in view of saving space.

In order to test $\mathcal{H}_0^{(2)}$, we first note that several independence or ‘‘portmanteau’’ tests for functional time series data have been recently proposed in the literature (e.g., Gabrys and Kokoszka, 2007; Zhang, 2016; Kokoszka et al.; 2017). Here we apply the portmanteau test proposed by Kokoszka et al. (2017). We consider the test statistic

$$V_{N,H} = N \sum_{h=1}^H \|\hat{\gamma}_{h,Z}\|^2$$

where H is a user specified maximum lag length, and the autocovariance function $\gamma_{h,Z}(t, s)$ is defined by,

$$\hat{\gamma}_{h,Z}(t, s) = \frac{1}{N} \sum_{i=1}^{N-h} [Z_i(t) - \bar{Z}(t)][Z_{i+h}(s) - \bar{Z}(s)].$$

According to Theorem 2 in Kokoszka et al. (2017), under $\mathcal{H}_0^{(2)}$

$$V_{N,H} \xrightarrow{d} \sum_{\ell=1}^{\infty} \xi_{\ell,H} \mathcal{N}_{\ell}^2 \stackrel{\mathcal{D}}{=} \Xi_{2,H}, \quad N \rightarrow \infty$$

where again \mathcal{N}_ℓ are standard normal random variables, and the $\xi'_{\ell,H}$ s are the eigenvalues of a covariance operator determined from the violation process. Kokoszka et al. (2017) describe approaches for estimating this limiting distribution, which we follow, and a P-value of the test is obtained as $p = P(\Xi_{2,H} > V_{N,H})$. Regarding the choice of H , several portmanteau-type backtesting studies indicate that selecting a large value of H tends to result in a loss of power to detect correlation in the violation process that tends to occur at short lags, (see, Escanciano and Olmo, 2010), and hence we consider only small and moderate values of $H = 1, 3, 5$ and 10 .

3. COMPARATIVE STUDY AND APPLICATION TO EQUITY AND FOREX MARKET RETURNS

In this section, we present the results of the empirical analysis that aims to evaluate each of the methods proposed above relative to benchmark methods for VaR forecasting at both daily and intra-day frequencies. The specific data that we consider consist of high resolution price data of three indices from equity markets and one FOREX market. The data series from equity markets are the S&P500, DAX30 and CAC40 indices, which are known as the leading indicators for business cycles in the U.S. and European markets. The foreign exchange data is the spot exchange rate between U.S. dollar (USD) and Euro. The sample consists of a five year period ranging from 2nd-January-2012 to 29th-March-2018, with a total of 1,566 trading days. The intra-day data is collected at the 1-minute frequency. Over the whole period, apart from the 2015-2016 stock market selloff, the global financial market is steadily rising over this period.

3.1. Model specification and benchmarks.

We compute the OCIDR curves according to Equation (1.2). The discretely sampled high-frequency returns are interpolated using the linear interpolation in order to produce full curves. We note that the linear interpolation can be readily justifiable in this setting relative to other potential interpolation methods, since if the prices evolve approximately as Brownian motions between the observed times, then the best predictor (conditional expectation) of the path given the observed endpoints should be linear.

Table 3.1 summarizes the trading hour and the total number of intra-day grid points for each asset, and provides basic summary statistics for the daily closing return $Y_i = X_i(1)$. One notable element given in this table is that the USD-Euro FOREX returns exhibit excess kurtosis

relative to the standard normal distribution, which suggests that the VaR forecasts assuming Gaussian innovations might not be appropriate.

TABLE 3.1. Summary information on the daily closing log returns of each of the series considered. No. Grid describes the number of one minute evaluation points of each OCIDR curve.

	Trading Hour (EST)	No. Grid	Mean	SD	Skew	Kurt	Min	Max
S&P500	9:30-16:00	390	0.0005	0.0077	-0.4943	3.1849	-0.0447	0.0381
DAX30	6:00-15:00	540	0.0004	0.0113	-0.3018	2.0684	-0.0687	0.0521
CAC40	6:00-15:00	540	0.0003	0.0113	-0.3426	3.1973	-0.0800	0.0450
USD-Euro	13:00-22:00	540	-0.0001	0.0054	0.6422	6.2241	-0.0216	0.0482

We compared the proposed methods to three families of benchmark models: the filtered historical simulation (FHS) based on the standard GARCH model, the realized GARCH (RGARCH) model (Hansen et al., 2012), and the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) model (Corsi, 2009). The first one is based on daily return information and is widely used in the banking sector; the latter two play more competitive roles as they model the daily conditional volatility by incorporating intra-day information. In particular the HAR model has been shown to be a strong benchmark in volatility forecasting.

The FHS method is described below. Similar to our two-step method, we fit Y_i with a GARCH(1,1) process first,

$$(3.1) \quad \begin{aligned} Y_i &= \sigma_i \varepsilon_i \\ \sigma_i^2 &= \omega + \alpha Y_{i-1}^2 + \beta \sigma_{i-1}^2, \end{aligned}$$

and estimate the scalar parameters by using QMLE, from which we obtain the fitted innovations $\hat{\varepsilon}_i$. We bootstrap $\hat{\varepsilon}_i$ and obtain the quantile corresponding to the empirical distribution $\hat{\varepsilon}_i^{B,\tau}$. For a one-step ahead forecasts this reduces to evaluating the sample τ th quantile of the residuals. The VaR is then computed as:

$$(3.2) \quad \widehat{\text{VaR}}_i^\tau = \hat{\sigma}_i \hat{\varepsilon}_i^{B,\tau}.$$

To describe the RGARCH benchmark, it is presumed that the process Y_i follows the first equation of (3.1), but the conditional volatility σ_i^2 follows a different recursive dynamic. In particular

$$(3.3) \quad \sigma_i^2 = \omega + \alpha RV_{i-1} + \beta \sigma_{i-1}^2,$$

where RV_i is a realized volatility modeled by the measurement equation $RV_i = \alpha + \beta\sigma_i^2 + \mathfrak{t}(\varepsilon_i) + \mathfrak{v}_i$. This measurement equation relates the realized volatility to the daily conditional volatility and leads to a joint modeling of volatility and asset returns. The term $\mathfrak{t}(\varepsilon_i)$ captures the leveraged effect and typically the \mathfrak{v}_i are taken to be IID normally distributed random variables. In the application, in order to conduct a fair comparison, we estimate the latent integrated volatility with an over-night effect by summing squared intra-day log returns using a quadratic variation estimate with the squared over-night return,

$$(3.4) \quad RV_i = \left[\sum_{j=2}^J (\log P_i(j \cdot \Delta) - \log P_i((j-1) \cdot \Delta))^2 + (\log P_i(0) - \log P_{i-1}(1))^2 \right]^{1/2},$$

where $P_i(j \cdot \Delta)$ is the discrete intra-day log return at day i with frequency Δ , and J is the number of intra-day log returns to be summed. The choice of Δ is crucial in defining the estimated quadratic variation volatility measure. Using the highest available frequency might cause this measure to be severely contaminated by market microstructure noises, see Meddahi (2002), while using a low frequency may lose useful information for measuring volatility. Here we take $\Delta = 5$ minutes. We also tried taking $\Delta = 1$ minute, but found the results to be generally worse. Model (3.3) is also estimated by using the QMLE method. Similarly, we calculate the VaR at level τ assuming either normal innovations or bootstrapped innovations,

$$\widehat{VaR}_i^\tau = \widehat{\sigma}_i \Phi^{-1}(\tau), \quad \widehat{VaR}_i^\tau = \widehat{\sigma}_i \widehat{\varepsilon}_i^{B,\tau}.$$

We call these benchmarks RGARCH-N and RGARCH-BOOT. Note that the VaR estimation according to different methods are notational indistinct.

Next we briefly review the HAR model. Seminal papers such as Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2003) showed that the integrated variance can be consistently approximated by the realized volatility. In this regard, a HAR model (Corsi, 2009) provides a parsimoniously specified dynamic model to forecast volatility among all types of realized volatility models. The HAR model directly works on the realized volatility (3.4), and the model can be specified as,

$$(3.5) \quad RV_i = \omega + \alpha RV_{i-1} + \beta RV_{i-1}^w + \gamma RV_{i-1}^m + u_i$$

where RV_i^w and RV_i^m are weekly and monthly realized volatility at time i , given as,

$$RV_i^w = 1/5(RV_i + \dots + RV_{i-4}), \quad RV_i^m = 1/22(RV_i + \dots + RV_{i-21})$$

Model (3.5) can be easily estimated by the standard LSE. We then calculate the VaR as,

$$\widehat{VaR}_i^\tau = \widehat{RV}_i \Phi^{-1}(\tau), \quad \widehat{VaR}_i^\tau = \widehat{RV}_i \widehat{\varepsilon}_i^{B,\tau}.$$

These benchmarks are termed as HAR-N and HAR-BOOT model, respectively.

We now describe the details of fitting the proposed models based on the OCIDR curves to the data sets under study. In order to estimate an FGARCH model for each of these series, we first estimate the FPCA basis derived from the squared processes $X_i^2(t)$. Note that return curves used for the FPCA are interpolated by using 30 cubic B-spline polynomials; see Chapter 3 in Ramsay and Silverman (2006) and Section 7 in Ramsay et al. (2009) for technical details. Figure 3.1 displays the first four estimated basis functions corresponding to the S&P500 OCIDR curves. The figure also shows how much variation is explained by each basis element. We notice that for all four series under study, the first two basis functions explain over 90% of the total variation. We thus take $M_1 = 2$ below. Similarly to selecting M_1 , we select M_2 which defines the number of projections in the GPD approximation of the residual distribution in order to explain 90% of the TVE. This results in taking $M_2 = 4$ in each data set.

FIGURE 3.1. First four PCA Basis elements of the squared process of S&P500 OCIDR curves.

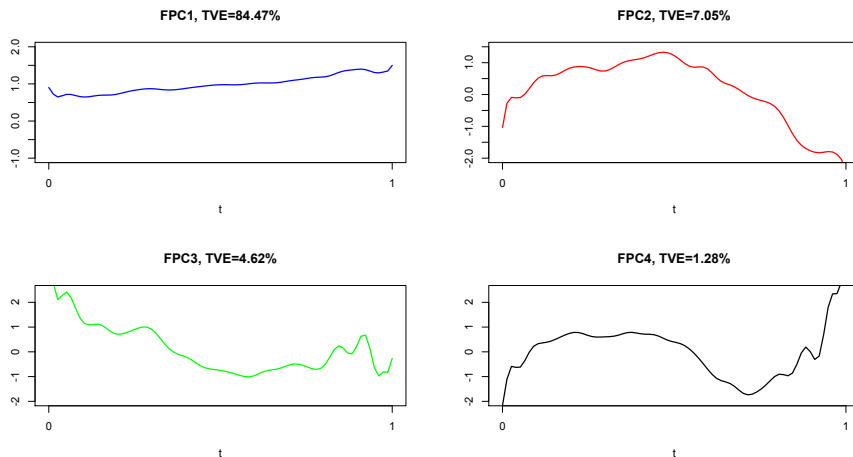


TABLE 3.2. P-values of goodness-of-fit tests of the FGARCH(1,1) model with $M_1 = 2$

	S&P500			DAX30			CAC40			USD-Euro		
	$H = 1$	$H = 5$	$H = 10$	$H = 1$	$H = 5$	$H = 10$	$H = 1$	$H = 5$	$H = 10$	$H = 1$	$H = 5$	$H = 10$
$V_{\varepsilon,N,H}$	0.45	0.90	0.97	0.37	0.91	0.94	0.96	1.00	1.00	0.15	0.31	0.71
$Q_{\varepsilon,N,H}$	0.33	0.73	0.96	0.26	0.75	0.84	0.98	1.00	1.00	0.16	0.40	0.78

Next, we evaluate the model adequacy of the FGARCH(p,q) candidates by using the diagnostic tests described above, starting from a FGARCH(1,1) specification suggested in Rice et al. (2019). Table 3.2 presents the goodness-of-fit test results in terms of approximate P-values for several choice of the maximum lag parameter H . The results for lag lengths $H = 1, 5$, and 10 indicate that the FGARCH(1,1) model is adequate for all of the series considered in this paper. As correctly pointed out by one of reviewers, here we select the parameters of the FGARCH model by using the entire sample. However, given our focus on forecasting, these three parameters could and perhaps should be selected recursively for each rolling window of the training sample. Aside from the goodness-of-fit testing applied above, one can then select the optimal model parameters through out-of-sample cross-validation. Here we mention two methods for the evaluation in the cross-validation step: backtesting the in-sample violation process, and assessing the forecasted quantiles through consistent scoring rules. We introduce these ideas in Section 4.2. In this paper though, considering the computational costs, we stick with our previous analysis and simply select the global optimal FGARCH model globally.

3.2. Empirical Results and Backtesting daily VaR Forecasts.

We forecast 1-day-ahead VaR for all of the assets based on a training sample of length 750 days (which is roughly the number of trading days in three years) and a testing sample of length 816 days. The forecasting method follows a rolling window approach, i.e., in every in-sample window, we use the past three years of observations to train the models including the FHS-GARCH(1,1), RGARCH(1,1), HAR and FGARCH(1,1) models, and then predict the 1-day-ahead VaR, rolling the window to forecast the out-of-sample spanning from 2nd-January-2015 to 29th-March-2018. In this exercise, we consider three quantiles: $\tau = 0.025, 0.01$, and 0.005.

In order to assess the performance of the forecasts of the risk measures, we evaluate the VaR from two perspectives, namely backtesting and capital requirement. The traditional ways to assess the validity of VaR have been to use the unconditional (Kupiec, 1995) and conditional

coverage tests (Christoffersen, 1998). There are other VaR backtesting methods available in the literature; see Nieto and Ruiz (2016) for a survey. Most of them concentrate on the violation process $V_i^\tau = 1(Y_i \leq \text{VaR}_i^\tau)$. Similar to the conditional coverage test, a valid risk measure should deliver an independent violation process with an unbiased expectation of τ , which is a testable hypothesis.

Table 3.3 documents the violation rates of the VaR forecasts for each method. The boxed values point out the method that produces the closest violation ratio to the true quantile τ . Apart from three values recommending the superiority of the parametric FG-EVT method, the other results show that the semi-parametric bootstrap-based methods performed the best in terms of the accuracy of violation ratios. In particular, the FG-B is tagged as the superior model in six out of twelve cases. Further, the $\widehat{\text{VaR}}_i^\tau$ forecasted assuming normal innovations shows the pattern that they are relatively accurate at the 2.5% quantile level with data from equity markets, but deviate when it comes to extreme quantiles.

TABLE 3.3. VaR violation rates with the boxed value indicating the model that produced the closest to nominal rate.

τ	S&P500			DAX30			CAC40			USD-Euro		
	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%
FHS	0.023	0.013	0.009	0.027	0.012	0.007	0.031	0.008	0.007	0.039	0.016	0.008
RG-N	0.028	0.020	0.016	0.029	0.016	0.011	0.029	0.019	0.015	0.044	0.029	0.026
RG-B	0.020	0.015	0.011	0.025	0.009	0.008	0.023	0.014	0.006	0.039	0.026	0.016
HAR-N	0.047	0.033	0.027	0.065	0.036	0.031	0.057	0.032	0.022	0.049	0.042	0.029
HAR-B	0.025	0.012	0.008	0.037	0.020	0.012	0.020	0.009	0.007	0.036	0.016	0.006
FG-OU	0.074	0.057	0.044	0.109	0.072	0.063	0.133	0.110	0.098	0.093	0.063	0.044
FG-B	0.021	0.011	0.008	0.019	0.009	0.003	0.024	0.007	0.004	0.034	0.018	0.010
FG-EVT	0.019	0.016	0.013	0.015	0.009	0.004	0.017	0.011	0.008	0.036	0.021	0.018

Table 3.4 reports results of the VaR backtesting, providing similar patterns of results with those reported in Table 3.3. Overall, the semi-parametric methods (FHS, RG-B, HAR-B, and FG-B) show better performance than the parametric methods (RG-N, HAR-N, FG-OU), with an exception that the FG-EVT model passes the backtests in ten out of twelve cases at the 5% level.

The results can be summarized as follows. First, the benchmark FHS, RGARCH, and HAR models forecast VaR reliably when the bootstrapped innovations are used, while the RGARCH method performs poorly in the case of USD-Euro. Similarly poor performances were observed to forecast the 1% and 0.5% quantiles, if the innovations are assumed to be normal. Additionally, the FGARCH model forecasts passed the backtests when the bootstrapped innovation were used,

particularly with equity data, but the performance is poor when normal innovations were used. Also, we notice that, as a fully parametric model, the FGARCH-EVT model performs relatively well in all cases.

In order to further compare the models, we also considered pair-wise and multi-model comparisons using the Diebold and Mariano test (DM test, Diebold and Mariano, 1995) and the model confidence set (MCS, Hansen et al., 2011) based on generalized probability scoring (PS); see Gneiting and Raftery (2007). The PS method can be used to compare quantile forecasts by using consistent scoring rules to evaluate the quality of the forecast based on the realized variable. Although many such consistent scores can be used, we consider the PS for the τ level quantile

$$(3.6) \quad S(f_{\mathcal{M}}(\tau), Y) = (f_{\mathcal{M}}(\tau) - Y)(\mathbb{1}_{\{f_{\mathcal{M}}(\tau) \leq Y\}} - \tau)$$

where $f_{\mathcal{M}}$ is the forecasting quantile of model \mathcal{M} , and Y is the realized outcome. Here a smaller score indicates a better forecast. In order to check the relative performances among all candidate models, we first calculate the averaged PS $\bar{S}_{\mathcal{M}}^f = \frac{1}{N-N_o+1} \sum_{i=N_o}^N S(f_{\mathcal{M}}(\tau), Y_i)$, where $\{Y_{N_o}, Y_{N_o+1}, \dots, Y_N\}$ are daily returns from the out-of-sample. Then, we compare scores pair-wise by using the DM test and calculating MCS. We omit the descriptions of these two tests to save space, but one can refer to the above papers for details.

TABLE 3.4. P-values of backtests for the VaR forecasts with the bold values indicating significance at the 5% level.

	S&P500			DAX30			CAC40			USD-Euro		
τ	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%
Panel A: Unconditional Coverage Test												
FHS	0.68	0.38	0.13	0.77	0.59	0.54	0.33	0.57	0.54	0.10	0.31	0.47
RG-N	0.60	0.02	0.00	0.46	0.13	0.06	0.46	0.03	0.00	0.03	0.00	0.00
RG-B	0.37	0.23	0.06	0.95	0.86	0.28	0.68	0.38	0.90	0.10	0.01	0.02
HAR-N	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HAR-B	0.98	0.59	0.28	0.04	0.02	0.02	0.37	0.86	0.54	0.18	0.31	0.95
FG-OU	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FG-B	0.39	0.19	0.54	0.39	0.92	0.61	0.63	0.60	0.91	0.36	0.31	0.41
FG-EVT	0.27	0.13	0.01	0.13	0.92	0.91	0.29	0.90	0.54	0.24	0.15	0.02
Panel B: Conditional Coverage Test												
FHS	0.64	0.20	0.05	0.81	0.78	0.80	0.59	0.81	0.80	0.23	0.54	0.75
RG-N	0.06	0.00	0.00	0.28	0.13	0.14	0.06	0.05	0.00	0.04	0.01	0.00
RG-B	0.01	0.17	0.15	0.23	0.92	0.54	0.15	0.20	0.97	0.08	0.02	0.01
HAR-N	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
HAR-B	0.80	0.78	0.54	0.13	0.04	0.06	0.49	0.92	0.80	0.24	0.54	0.99
FG-OU	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FG-B	0.37	0.85	0.28	0.25	0.86	0.32	0.86	0.33	0.69	0.29	0.15	0.19
FG-EVT	0.25	0.13	0.01	0.05	0.86	0.69	0.16	0.85	0.28	0.18	0.06	0.00

Table 3.5 reports the average PS for all models. Similarly as in the above analyses, the bootstrap-based models generally performed better. The FHS, RG-B and HAR-B models are the best model in most cases, but the FG-B and FG-EVT models have comparable scores.

TABLE 3.5. Average probability scores of the VaR forecasts with the boxed value indicating the model that produces smallest scores, all scores are valued with units 10^{-4} .

	S&P500			DAX30			CAC40			USD-Euro		
τ	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%
FHS	5.72	2.82	1.69	7.74	3.73	<u>2.01</u>	<u>8.16</u>	<u>4.24</u>	2.50	<u>4.16</u>	<u>1.95</u>	<u>1.18</u>
RG-N	5.72	3.16	2.08	7.78	3.93	2.37	8.20	4.37	2.77	5.69	2.79	1.59
RG-B	5.65	2.99	1.94	<u>7.69</u>	<u>3.68</u>	2.10	8.21	4.26	<u>2.46</u>	4.29	2.17	1.21
HAR-N	5.92	3.45	2.36	8.96	4.89	3.25	8.74	4.92	3.35	4.36	2.28	1.37
HAR-B	<u>5.57</u>	<u>2.76</u>	<u>1.61</u>	8.23	4.10	2.58	8.26	4.32	2.63	4.18	2.03	1.15
FG-OU	7.10	4.68	3.56	10.71	6.67	4.83	11.97	8.38	6.60	5.21	3.10	2.24
FG-B	6.19	3.22	1.90	8.45	4.03	2.15	8.74	4.48	2.47	4.54	2.23	1.26
FG-EVT	6.11	3.10	2.03	8.69	3.93	2.08	8.93	4.29	2.49	4.56	2.30	1.40

Table 3.6 displays the results of testing the significance of these differences in performance.

Panel A selects the “better” models under 25% model confidence set. It is notable that the

FG-OU model are eliminated in almost all cases, as well as the HAR-N model with equity data at extreme quantiles. Given these results, in Panel B, we exclude the FG-OU model, and pairwise compare the rest of models by using the DM test. The testing results basically reinforce the findings mentioned above, i.e., the bootstrap-based models significantly improve over the normal distribution based models. Also, all bootstrap-based models were not significantly different (at the 0.05 level) in pairwise comparison, including the FG-B model, the exception being the RG-B method with the FOREX index. The FG-EVT model performed statistically indistinguishable from to the bootstrap-based models.

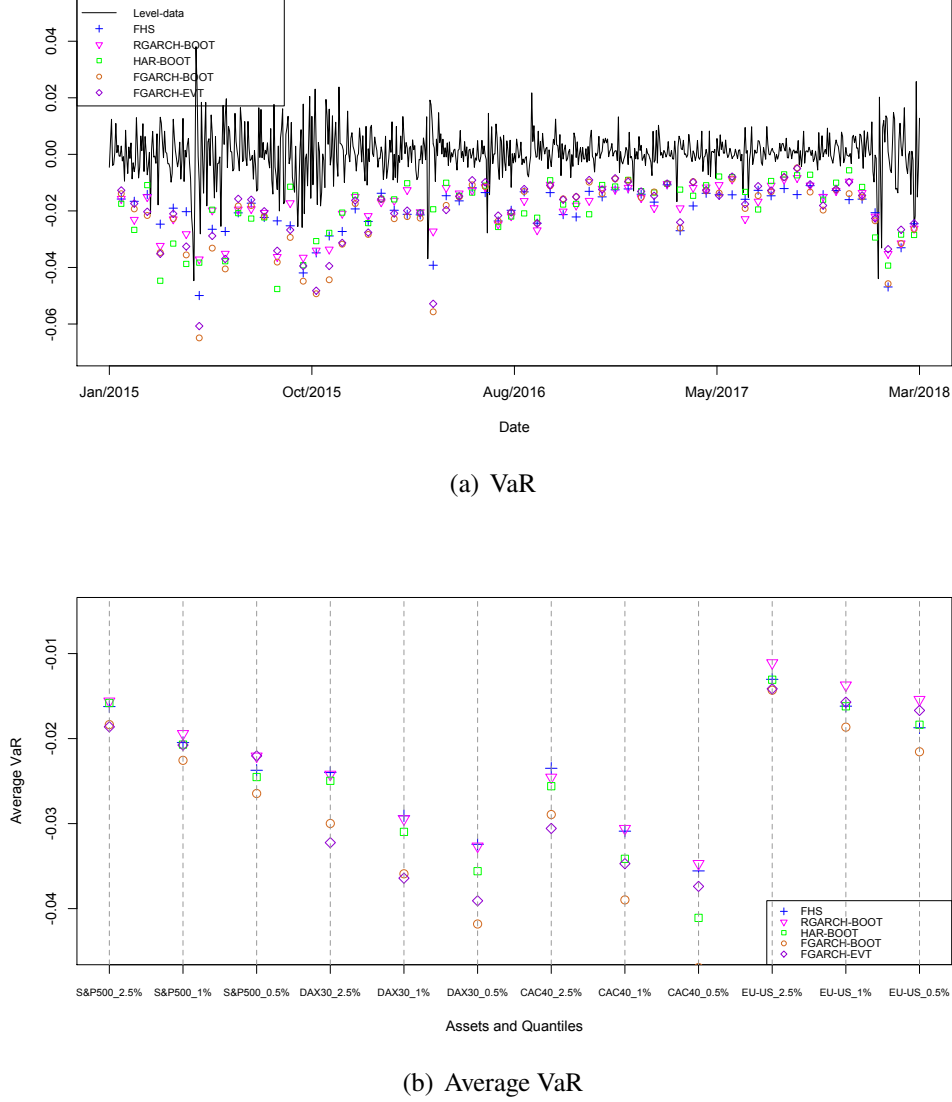
Finally, we also examined the cap requirements each model imposes. Part (a) in Figure 3.2 displays an example plot of the VaR predictions for S&P500 with quantile $\tau = 0.01$ for those methods that performs well in backtests. From the plot, we can see that the market is relatively stable at the beginning of 2015, while it is relatively more turbulent during the 2015-2016 stock market selloff. We also observe the benchmarks (FHS, RGARCH-BOOT, HAR-BOOT) forecast VaR to be relatively tight, while we observe that the FGARCH-based predictors to relatively deviate from the data. This pattern is consistent during both the period of an economic expansion and the period of a market turbulence.

TABLE 3.6. Results of relevant performance tests of the VaR forecasts. In Panel A, the bold values mean those models belong to the 25% MCS. Panel B contains the DM test statistics, with the asterisk *, **, *** denoting tests at 10%, 5% and 1% significance levels, respectively.

	S&P500			DAX30			CAC40			USD-Euro		
	2.5%	1.0%	0.5%	2.5%	1.0%	0.5%	2.5%	1.0%	0.5%	2.5%	1.0%	0.5%
Panel A: P values of Model Confidence Set												
FHS	0.96	1.00	0.90	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RG-N	0.99	0.70	0.29	0.90	0.70	0.29	1.00	1.00	0.68	0.64	0.50	0.76
RG-B	0.99	0.75	0.29	1.00	1.00	0.45	1.00	1.00	1.00	0.99	0.81	1.00
HAR-N	0.73	-	-	-	-	-	0.27	0.33	-	0.97	0.81	0.82
HAR-B	1.00	1.00	1.00	-	0.37	-	1.00	1.00	0.97	1.00	1.00	1.00
FG-OU	-	-	-	-	-	-	-	-	-	-	-	-
FG-B	-	-	-	-	-	0.70	-	0.91	1.00	0.46	0.61	0.97
FG-EVT	-	0.30	-	-	0.67	0.97	-	1.00	1.00	0.47	0.59	0.66
Panel B: Diebold & Mariano Test												
FHS v RG-N	-1.39	-1.56	-1.62	-2.10**	-2.01**	-1.64	-1.52	-1.37	-1.41	-1.54	-1.79*	-1.88*
FHS v RG-B	-0.70	-0.92	-1.41	0.58	0.81	0.63	1.21	-0.34	-0.89	-2.60***	-1.81*	-1.25
FHS v HAR-N	-2.09**	-1.99**	-1.99**	-2.90***	-2.72***	-2.39**	-2.18**	-1.75*	-1.77*	-1.03	-1.18	-0.81
FHS v HAR-B	-0.08	0.67	0.93	0.46	0.34	-1.94*	0.96	1.12	0.96	0.11	-0.24	0.20
FHS v FG-B	-0.61	-0.80	-0.65	1.21	1.03	0.98	1.53	1.64	1.17	-0.16	0.42	0.57
FHS v FG-EVT	0.55	-0.89	-1.59	1.23	1.08	0.99	1.45	1.29	0.90	-0.45	-1.01	-0.87
RG-N v RG-B	1.59	2.18**	1.71*	2.40**	1.65	1.51	1.90*	1.41	1.52	1.41	1.42	1.64
RG-N v HAR-N	-2.16**	-1.79*	-1.52	-2.06**	-1.95*	-1.94*	-1.26	-0.98	-0.84	1.39	1.17	0.60
RG-N v HAR-B	1.71*	1.87*	1.79*	1.15	1.08	0.94	1.47	1.31	1.32	1.55	1.58	1.48
RG-N v FG-B	1.28	1.45	1.37	1.65*	1.43	1.27	1.59	1.51	1.34	1.53	1.79*	1.70*
RG-N v FG-EVT	1.61	1.45	1.13	1.61	1.43	1.29	1.53	1.36	1.30	1.51	1.48	1.16
RG-B v HAR-N	-3.10***	-3.15***	-2.56**	-2.97***	-2.35**	-2.26**	-2.44**	-1.69*	-1.87	-0.01	-0.22	-0.65
RG-B v HAR-B	1.48	1.57	1.64	0.37	-0.22	-1.80*	0.51	1.08	1.01	1.08	0.61	0.56
RG-B v FG-B	0.56	0.53	0.97	1.27	1.13	1.03	1.29	1.49	1.16	1.34	1.13	0.92
RG-B v FG-EVT	1.06	0.45	0.10	1.28	1.19	1.04	1.23	1.27	0.97	1.14	0.36	-0.50
HAR-N v HAR-B	2.80***	2.36**	2.18**	2.41**	2.11**	2.41**	2.47**	1.68*	1.61	2.82***	2.52***	2.00**
HAR-N v FG-B	2.35**	2.07**	1.78*	2.38**	2.06**	1.84*	2.07**	1.86*	1.55	1.69*	1.99**	1.74*
HAR-N v FG-EVT	2.35**	2.09**	1.77*	2.30**	2.04**	1.95*	1.88*	1.64	1.55	1.42	1.18	0.71
HAR-B v FG-B	-1.04	-1.82*	-1.52	1.83*	1.71*	1.34	1.13	1.01	1.28	-0.43	1.07	0.88
HAR-B v FG-EVT	0.49	-1.77*	-2.04**	1.88*	1.77*	1.37	1.04	0.55	0.16	-0.68	-1.12	-1.66*
FG-B v FG-EVT	1.29	-0.77	-1.64	1.23	1.45	-0.89	0.57	-0.17	-1.48	-0.96	-2.15**	-1.88*

Then, we plot the average capital reserve required by the VaR for $\tau = 2.5\%$, 1% and 0.5% for both of the equity and FOREX markets. We make three remarks in part (b). First, the benchmarks (FHS and HAR-BEET) perform reasonably well and show a moderate level of capital requirement for all of the assets considered in this study at the chosen quantile levels. Second, the RGARCH-BEET model delivers relatively smaller expected losses and provides the riskiest measurement. Lastly, at the majority of the selected quantile levels, the FGARCH-BEET and FGARCH-EVT methods generally produce the most conservative VaR measures. In the Appendix, we also provide the Expected Shortfall forecasts, which are obtained based on the VaR forecasts, and the backtesting and relative accuracy hypotheses are tested and discussed as well.

FIGURE 3.2. (Top panel) Plots of VaR forecasts at $\tau = 0.01$ for S&P500. The forecasts are plotted every twelfth day in order to aid in differentiation of the models. (Bottom panel) The average VaR for assets at $\tau = 2.5\%$, 1% , 0.5% .



3.3. Empirical Results and Backtesting Intra-day VaR Curves.

In this last subsection, we provide an application of forecasting and backtesting intra-day VaR curves by using the methods proposed in Section 2.2. Figure 3.3 illustrates the predicted intraday VaR curves (recall eq. (2.2)) for the index S&P500 between 16th-March and 29th-March 2018 through the FGARCH(1,1) model with three types of innovations. It shows similar patterns with the summary drawn from the previous analyses at the daily frequency, i.e., the FGARCH-BQOT generates the most conservative intra-day VaR curves, and the VaR curves with the OU

innovation are more tight to the data but more likely to produce violations. The FGARCH-EVT curves lie generally between the two.

FIGURE 3.3. Plot of 10 Days of 1-day ahead VaR Curve forecasts at $\tau = 0.01$ for S&P500 in 2018.

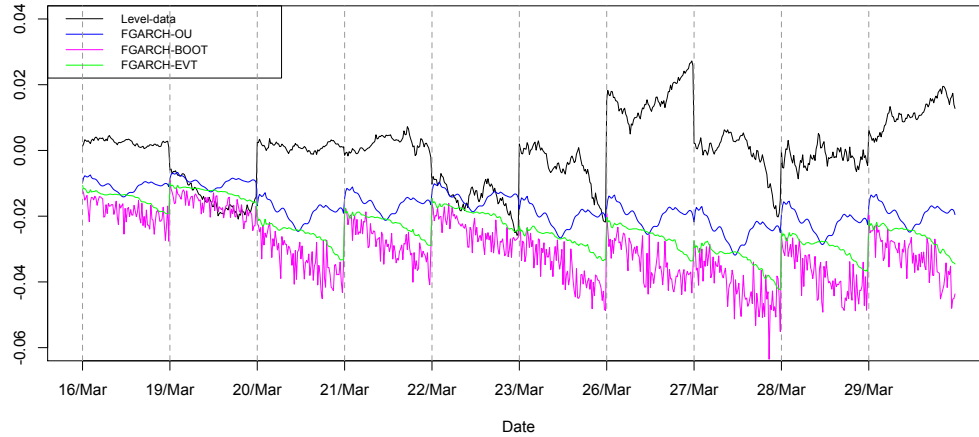


TABLE 3.7. P-values of backtests of \mathcal{H}_0^1 and \mathcal{H}_0^2 for the VaR curves forecasts with the bold values indicating the significance at the 5% Level

	S&P500			DAX30			CAC40			USD-Euro		
	FG-OU	FG-B	FG-EVT	FG-OU	FG-B	FG-EVT	FG-OU	FG-B	FG-EVT	FG-OU	FG-B	FG-EVT
$\mathcal{H}_0^{(1)}$												
2.5%	0.00	0.32	0.44	0.00	0.47	0.03	0.00	0.49	0.53	0.00	0.38	0.36
1.0%	0.00	0.39	0.00	0.00	0.49	0.06	0.00	0.45	0.00	0.00	0.38	0.10
0.5%	0.00	0.16	0.00	0.00	0.47	0.00	0.00	0.36	0.00	0.00	0.17	0.00
$\mathcal{H}_0^{(2)}, H=1$												
2.5%	0.51	0.22	0.18	0.69	0.51	0.35	0.13	0.22	0.33	0.08	0.29	0.27
1.0%	0.44	0.16	0.23	0.60	0.00	0.06	0.21	0.00	0.15	0.00	0.05	0.03
0.5%	0.36	0.03	0.01	0.68	0.03	0.01	0.17	0.05	0.12	0.00	0.14	0.05
$\mathcal{H}_0^{(2)}, H=3$												
2.5%	0.75	0.28	0.28	0.80	0.53	0.44	0.35	0.26	0.26	0.45	0.33	0.32
1.0%	0.71	0.00	0.14	0.86	0.23	0.31	0.37	0.09	0.20	0.39	0.07	0.05
0.5%	0.66	0.05	0.00	0.82	0.01	0.31	0.24	0.01	0.01	0.22	0.16	0.20
$\mathcal{H}_0^{(2)}, H=5$												
2.5%	0.47	0.27	0.26	0.53	0.29	0.29	0.48	0.10	0.01	0.52	0.44	0.48
1.0%	0.51	0.30	0.25	0.83	0.23	0.22	0.43	0.00	0.08	0.62	0.32	0.46
0.5%	0.42	0.32	0.27	0.76	0.00	0.22	0.30	0.01	0.00	0.57	0.27	0.38
$\mathcal{H}_0^{(2)}, H=10$												
2.5%	0.21	0.19	0.16	0.14	0.31	0.26	0.58	0.25	0.24	0.51	0.53	0.48
1.0%	0.15	0.26	0.15	0.39	0.34	0.26	0.46	0.31	0.30	0.51	0.39	0.45
0.5%	0.09	0.26	0.13	0.54	0.01	0.26	0.26	0.33	0.30	0.74	0.32	0.41

Table 3.7 reports the P-values of backtesting the conditional coverage and independence hypotheses on the violation functions defined by (2.12). We can see that FG-OU model cannot pass the conditional coverage hypothesis, and the same results occur in FG-EVT in forecasting extreme VaR curves. The FG-B model, as the one that passes $\mathcal{H}_0^{(1)}$ in all cases, generally

produces independent violation functions across different values of H , though there are a few exceptions in the index of CAC40. The FG-EVT appears to be generally valid to forecast intra-day VaR curves if a less extreme quantile is considered.

4. CONCLUSION

We considered the issue of forecasting daily and intra-day Value-at-Risk by using fine-resolution intra-day trading information. In order to capture the information of intra-day price movement and the over-night effect, we introduced and studied the over-night cumulative intra-day return curves, and their incorporation into risk modeling using a functional time series framework. Several forecasting methods were discussed, including parametric and semi-parametric two-step approaches. These are mainly based on the functional GARCH(p,q) model. Exploiting the structure of the FGARCH model, we also proposed a new approach to forecast intra-day VaR curves, and corresponding backtests for such curves.

In the application, we considered three equity indices (S&P500, DAX30, and CAC40) and one FOREX spot rate (USD-Euro) at the 1-minute frequency. After smoothing these high-frequency data into daily curves, we forecast the one-day-ahead VaR by using the proposed methods, along with three benchmark models – the filtered historical simulation based on daily information, the realized GARCH, and the heterogeneous autoregressive model based on the intra-day realized measurement. The empirical findings indicated that our methods based on the FGARCH model with bootstrapped residuals yielded comparatively good performance to VaR forecasts over benchmark models. The FGARCH-EVT model also was shown to provide reasonable forecasts in the equity markets and FOREX markets. Regarding the intra-day VaR forecasting, the FGARCH-BOOT model produced generally valid VaR curves, meanwhile, the good performances are shown in the FGARCH-EVT model to forecast less extreme VaR curves. It is worth mentioning that our methodologies can be generally applied to study the tail behavior of other arbitrarily fine grid observed economic and financial data. Avenues of future researches in this direction might focus on incorporating the leverage effect for intra-day return curves from the equity market, and to augment the quantile regression model with an explicit autoregressive term based on the conditional quantiles.

5. ACKNOWLEDGMENTS

This paper is dedicated to the memory of the late Peter Christoffersen, who passed away on June 22, 2018. Peter, among his many and enduring contributions to the profession, was the author of the book: *Elements of Financial Risk Management*. His work on VaR estimation and backtesting influenced the authors greatly in preparing this paper. The authors would also like to thank the editor, Esther Ruiz, the associate editor, and two anonymous referees, whose detailed and constructive comments and suggestions have helped improve both the content and presentation of this paper. One referee in particular deserves special recognition for their help in improving nearly every aspect of our paper over the course of the review process—their efforts went well beyond the normal role of a referee. The usual disclaimer applies.

REFERENCES

- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica* 71, 529–626.
- Angelidis, T., Degiannakis, S. (2008). Volatility forecasting: Intra-day versus inter-day models. *Journal of International Financial Markets, Institutions and Money* 18, 449–465.
- Aue, A., Horváth, L., Pellatt, D., 2017. Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis* 38, 3–21.
- Barndorff-Nielsen, O., Shephard, N., 2002. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* 17, 457–477.
- Basel Committee (1996). Overview of the Amendment of the Capital Accord to Incorporate Market Risk. Basel Committee on Banking Supervision.
- Basel Committee (2010). Group of Governors and Heads of Supervision announces higher global minimum capital standards. Basel Committee on Banking Supervision.
- Bosq, D. (2000) *Linear Processes in Function Spaces: Theory and Applications*. Springer.
- Ljung, G. M., Box, G. E. P. (1978). On a measure of a lack of fit in time series models. *Biometrika*. 65, 297–303.
- Cerovecki, C., Francq, C., Hörmann, S., Zakoian, J. (2019). Functional GARCH models: the quasi-likelihood approach and its applications. *Journal of Econometrics* .
- Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics* 33, 806–839.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review* 39, 841–862.
- Christoffersen, P. (2010). Backtesting. *Encyclopedia of Quantitative Finance*, Wiley.
- Christoffersen, P. (2011). *Elements of financial risk management*. Academic Press.

- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7, 174–196.
- Diebold, F. X., Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20, 134–144.
- Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.
- Engle, R. F., Manganelli, S. (2004). CAViaR: conditional autoregressive Value at Risk by regression quantiles. *Journal of Business & Economics Statistics* 22, 367–381.
- Escanciano, J. C., Olmo, J. (2010). Backtesting parametric value-at-risk with estimation risk. *Journal of Business & Economic Statistics*, 28, 36–51.
- Ferraty, F., Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer.
- Ferraty, F., Quintela-del-Río, A. (2016). Conditional VaR and expected shortfall: a new functional approach. *Econometric reviews* 35, 263–292.
- Fuest, A., Stefan, M. (2015). Modeling liquidity impact on volatility: a GARCH-FunXL approach. *SSRN Library*.
- Fink, H., Fuest, A., Port, H. (2018). The impact of sovereign yield curve differentials on Value-at-Risk forecasts for foreign exchange rates. *Risks*.
- Francq, C., Zakoian, J. M. (2011). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.
- Gabrys, R., Kokoszka, P. (2007). Portmanteau test of independence for functional observations. *Journal of the American Statistical Association* 102, 1338–1348.
- 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.
- Gerlach, R., Wang, C. (2016). Forecasting risk via realized GARCH, incorporating the realized range. *Quantitative Finance* 16, 501–511.
- Giot, P., Laurent, S. (2004). Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance* 11, 379–398.
- Gneiting, T., Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102, 359–378.
- Gomes, M. I., Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: a review. *International Statistical Review* 83, 263–292.
- Hansen, P. R., Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* 24, 127–161.
- Hansen, P. R., Lunde, A., Nason, J. M. (2011). The model confidence set. *Econometrica*, 79, 453–497.

- 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.
- Hörmann, S., Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics* 38, 1845–1884.
- Horta, E., Ziegelmann, F. (2018). Dynamics of financial returns densities: A functional approach applied to the Bovespa intraday index. *International Journal of Forecasting* 34, 75–88.
- Horváth, L., Kokoszka, P. (2012). *Inference for functional data with applications*. Springer.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* 3.
- Kokoszka, P., Reimherr, M. (2013). Predictability of shapes of intraday price curves. *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.
- Kokoszka, P., Xiong, Q. (2018). Extremes of projections of functional time series on data-driven basis systems. *Extremes* 21, 177–204.
- Kokoszka, P., Zhang, X. (2012). Functional prediction of cumulative intraday returns. *Statistical Modelling* 12, 377–398.
- McNeil, A. J., Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7, 271–300.
- Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics* 17, 479–508.
- Meng, X., Taylor, J. W. (2018). An approximate long-memory range-based approach for value at risk estimation. *International Journal of Forecasting* 34, 377–388.
- Nieto, M. R., Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting* 32, 475–501.
- Ramsay, J.O., Silverman, B.W. (2006). *Functional Data Analysis*. Wiley Online Library.
- Ramsay, J. O., Hooker, G., Graves, S. (2009). *Functional Data Analysis with R and MATLAB. (Use R)*, Springer.
- Rice, G., Wirjanto, T., Zhao, Y. (2019). Tests for conditional heteroscedasticity with functional data and goodness-of-fit tests for FGARCH models. MPRA Paper.
- Slovik, P., Cournède, B. (2011). Macroeconomic impact of Basel III. *OECD Economics Department Working Papers. No. 844, OECD Publishing, Paris*.
- Zhang, X. (2016). White noise testing and model diagnostic checking for functional time series. *Journal of Econometrics* 194, 76–95.

APPENDIX A. FORECASTING EXPECTED SHORTFALL

In this appendix, we discuss how to obtain a simple forecast of the Expected Shortfall (ES) based on the above VaR forecasts. The ES at quantile τ is usually defined, in the low-frequency literature, as

$$\text{ES}_i^\tau = \mathbb{E}[Y_i | \mathcal{F}_{i-1}, Y_i < \text{VaR}_i^\tau].$$

The analytic solutions for various ES estimators at daily frequency are reviewed by Nadarajah et al. (2013). To adapt the definition of ES into the functional high-frequency context we obtain,

$$\text{ES}_i^\tau(t) = \mathbb{E}[X_i(t) | \mathcal{F}_{i-1}, X_i(t) < \text{VaR}_i^\tau(t)]$$

Here we directly estimate the intra-day ES of $X_i(t)$ from the sample as,

$$(A.1) \quad \widehat{\text{ES}}_i^\tau(t) = \frac{1}{N(t)} \sum_{j=1}^N X_i(t) \mathbb{1}_{\{X_i(t) < \widehat{\text{VaR}}_i^\tau(t)\}},$$

where $N(t) = \sum_{j=1}^N \mathbb{1}_{\{X_i(t) < \widehat{\text{VaR}}_i^\tau(t)\}}$. Thus the one-day-ahead ES curve is predicted by taking the pointwise sample mean of the observations that exceed $\widehat{\text{VaR}}_i^\tau(t)$. As such, $\widehat{\text{ES}}_i^\tau(1)$ gives an estimate of the daily ES.

To backtest daily ES, we focus on the event of quantile violations and apply the daily standardized residual test used in McNeil and Frey (2000). Let $(Y_i < \text{VaR}_i^\tau)$ denote those daily returns hit the VaR forecasts, and then the daily standardized residuals r_i can be computed by,

$$r_i = \frac{(Y_i < \text{VaR}_i^\tau) - \text{ES}_i^\tau}{\hat{\sigma}_i},$$

where $\hat{\sigma}_i$ is the estimated conditional standard error. Another possible way to estimate the conditional density function is to use the method proposed in Fernandes et al. (2019). It is clear that r_i focuses on the difference between Y_i and ES_i^τ in the event of violations. We aim to test one-sided hypotheses on $\mathbb{E}[r_i]$,

$$\mathcal{H}_0^E: \mathbb{E}[r_i] \geq 0 \text{ against } \mathcal{H}_A^E: \mathbb{E}[r_i] < 0.$$

The one-sided backtest is appealing because the overestimation of the ES results in a failure of risk management and a punishment by the regularity authority, while the underestimation of the ES does not cause this problem. Thus, the ES is deemed to be acceptable if we cannot reject \mathcal{H}_0^E .

TABLE A.1. P-values of Backtests of \mathcal{H}_0^E for the ES Forecasts with the Bold Values Indicating the Significance at 95% Level.

	S&P500			DAX30			CAC40			USD-Euro		
τ	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%
FHS	0.05	0.53	0.53	0.55	0.29	0.55	0.55	0.07	0.09	0.59	0.59	0.23
RG-N	0.16	0.10	0.18	0.28	0.06	0.03	0.24	0.25	0.29	0.04	0.24	0.64
RG-B	0.07	0.16	0.06	0.51	0.18	0.56	0.36	0.31	0.06	0.02	0.05	0.63
HAR-N	0.33	0.54	0.53	0.73	0.65	0.68	0.73	0.36	0.09	0.10	0.41	0.38
HAR-B	0.40	0.76	0.69	0.87	0.80	0.75	0.05	0.26	0.63	0.39	0.43	0.00
FG-OU	0.53	0.61	0.28	0.92	0.83	0.86	0.91	0.91	0.93	0.37	0.24	0.10
FG-B	0.04	0.08	0.57	0.75	0.37	0.00	0.74	0.09	0.76	0.39	0.42	0.35
FG-EVT	0.09	0.27	0.14	0.42	0.55	0.74	0.55	0.11	0.21	0.54	0.07	0.09

Table A.1 exhibits the backtesting results. The majority of models provide reasonable ES, apart from a few exceptions in the RGARCH model and bootstrap-typed methods in HAR and FGARCH models.

In order to further assess the relative accuracy of the above models, we calculate the PS for ES forecasts according to Equation (4.6). Table A.2 shows the average PS, and similar to Table 4.5, the FGARCH model provides comparable PS with bench-marks. However, the models with normal distributed innovation generally outperform bootstrap-typed models, particularly at less extreme quantiles. We attribute this to the reason that bootstrapped models are more likely to provide much conservative VaR forecasts.

TABLE A.2. Average Probability Scores of the ES Forecasts with the Boxed value Indicating the Model that Produces Smallest Scores, all Scores are Valued with Unit e-4.

	S&P500			DAX30			CAC40			USD-Euro		
τ	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%	2.5%	1%	0.5%
FHS	6.73	3.31	2.01	8.68	4.05	2.39	9.38	4.59	2.73	4.35	1.96	1.11
RG-N	6.16	3.19	2.01	8.56	4.09	2.39	8.87	4.44	2.66	4.14	1.93	1.06
RG-B	6.51	3.31	2.16	8.73	4.10	2.42	9.39	4.66	2.79	4.21	1.97	1.04
HAR-N	5.83	3.04	1.89	8.32	4.15	2.50	8.70	4.53	2.85	4.33	2.27	1.32
HAR-B	6.32	3.24	1.99	8.44	4.13	2.58	9.35	4.69	2.89	4.41	2.03	1.16
FG-OU	6.07	3.38	2.36	8.08	4.13	2.55	8.16	4.43	2.93	4.12	2.22	1.47
FG-B	6.96	3.45	2.18	9.08	4.36	2.46	9.62	4.84	3.00	4.29	2.02	1.17
FG-EVT	6.96	3.17	2.08	9.50	4.31	2.52	9.95	4.59	2.69	4.35	2.15	1.15

We then apply the Model Confidence Set and Diebold & Mariano test to evaluate the relative performances of these models. In the panel A of Table A.3, we can see that the HAR-N and FG-OU models are belonging to the 25% model confidence set in most of cases. Panel B shows the pair-wise comparisons of ES forecasts by using the Diebold & Mariano test. The results consistently match our summaries made above, i.e., the majority of the models are statistically equivalent, with few cases indicating that HAR-N model performs better.

TABLE A.3. Results of Relevant Performance Tests of the ES Forecasts. In Panel A, the bold values mean that models are belong to the 25% MCS. In Panel B, the asterisk *, **, *** mean significant at 90%, 95% and 99% significance levels.

	S&P500			DAX30			CAC40			USD-Euro		
	2.5%	1.0%	0.5%	2.5%	1.0%	0.5%	2.5%	1.0%	0.5%	2.5%	1.0%	0.5%
Panel A: P values of Model Confidence Set												
FHS	-	0.81	0.97	0.63	1.00	1.00	-	1.00	0.89	0.70	1.00	-
RG-N	-	-	-	0.57	1.00	1.00	-	1.00	1.00	1.00	1.00	0.89
RG-B	-	-	0.42	0.31	1.00	1.00	-	0.97	-	0.38	0.78	1.00
HAR-N	1.00	1.00	1.00	0.88	1.00	0.99	-	1.00	0.68	0.96	0.52	0.40
HAR-B	-	0.95	0.99	0.94	1.00	0.37	-	0.95	-	0.78	0.94	0.48
FG-OU	0.37	0.48	-	1.00	1.00	0.98	1.00	1.00	0.77	1.00	0.87	-
FG-B	-	0.26	-	-	0.34	0.50	-	-	-	0.93	0.60	-
FG-EVT	-	0.98	0.82	-	-	-	-	1.00	1.00	0.77	0.72	0.55
Panel B: Diebold & Mariano Test												
FHS v RG-N	0.39	-0.33	-1.24	-1.77*	-1.27	-0.48	-0.95	-1.11	-1.17	-1.25	-1.14	-1.00
FHS v RG-B	0.43	0.12	-1.70*	0.72	0.95	0.98	1.48	-0.32	0.76	-1.63	-1.09	9.55***
FHS v HAR-N	0.36	-0.69	-0.83	-2.17**	-2.11**	-1.13	-1.02	-1.42	-1.41	-1.81*	-1.72	-1.45
FHS v HAR-B	0.92	0.27	-0.86	0.78	0.56	-1.41	1.02	-0.83	-0.62	-1.63	-1.32	-1.14
FHS v FG-B	0.26	0.02	-1.05	0.86	-0.19	0.89	1.23	-0.62	-1.02	0.78	-1.02	-1.13
FHS v FG-EVT	1.06	0.97	-1.54	0.29	0.16	-1.12	1.18	0.93	-0.82	0.69	-1.27	-1.15
RG-N v RG-B	0.05	0.86	-0.28	1.61	1.09	1.00	1.22	1.09	1.19	-1.71*	0.47	1.04
RG-N v HAR-N	-0.09	-1.12	0.71	-1.86*	-1.94*	-1.46	-0.12	-1.49	-1.72*	-1.66*	-1.94	-1.55
RG-N v HAR-B	1.21	0.39	0.38	1.19	0.86	0.07	1.01	0.98	1.14	-1.25	-1.23	-1.23
RG-N v FG-B	-0.60	0.31	0.81	1.17	0.27	0.96	1.15	1.05	0.63	1.01	0.84	-1.29
RG-N v FG-EVT	0.35	1.11	0.11	0.81	0.57	0.28	1.12	1.16	1.08	0.93	-1.34	-1.24
RG-B v HAR-N	-0.07	-1.83*	0.49	-2.21	-1.76*	-1.61	-1.26	-1.43	-1.43	-1.55	-1.97**	-1.46
RG-B v HAR-B	1.55	0.07	0.87	0.79	-1.24	-1.61	0.47	-1.18	-1.12	-0.98	-1.36	-1.16
RG-B v FG-B	-0.67	-0.11	0.96	0.88	-1.12	-1.16	1.10	-1.06	-1.06	1.16	0.75	-1.15
RG-B v FG-EVT	0.35	0.95	0.43	0.16	-1.04	-1.01	1.04	1.07	-0.96	1.09	-1.39	-1.16
HAR-N v HAR-B	1.16	0.72	0.12	1.73*	1.66	0.80	1.05	1.36	1.39	1.98**	2.10**	1.70*
HAR-N v FG-B	-0.51	0.67	0.21	1.58	1.02	1.58	1.20	1.40	1.15	1.54	2.01**	1.58
HAR-N v FG-EVT	0.36	1.45	-0.15	1.33	1.34	0.92	1.17	1.47	1.38	1.50	2.10**	1.70*
HAR-B v FG-B	-2.15**	-0.28	-0.06	0.81	-0.75	1.59	1.35	1.22	-0.96	1.30	1.57	1.16
HAR-B v FG-EVT	-0.34	0.68	-0.58	-0.85	-0.49	0.68	1.26	1.15	-0.58	1.25	-1.07	-0.60
FG-B v FG-EVT	1.01	1.33	-0.42	-1.39	1.26	-0.95	-1.41	1.09	1.09	-0.99	-1.47	-1.19

REFERENCES

- Fernandes, M.; Guerre, E.; Horta, E. (2019). Smoothing quantile regressions. *Journal of Business & Economic Statistics*, DOI: 10.1080/07350015.2019.1660177.
- Nadarajah, S., Zhang, B, Chan, S. (2013). Estimation methods for expected shortfall. *Quantitative Finance* 14, 271–291.
- McNeil, A. J., Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7, 271–300.