Econometric Theory, **29**, 2013, 267–288.
doi:10.1017/S0266466612000345

A FUNCTIONAL VERSION OF THE ARCH MODEL

SIEGFRIED HÖRMANN
Université Libre de Bruxelles

LAJOS HORVÁTH AND RON REEDER
University of Utah

Improvements in data acquisition and processing techniques have led to an almost continuous flow of information for financial data. High-resolution tick data are available and can be quite conveniently described by a continuous-time process. It is therefore natural to ask for possible extensions of financial time series models to a functional setup. In this paper we propose a functional version of the popular autoregressive conditional heteroskedasticity model. We will establish conditions for the existence of a strictly stationary solution, derive weak dependence and moment conditions, show consistency of the estimators, and perform a small empirical study demonstrating how our model matches with real data.

1. INTRODUCTION

To date, not many functional time series models exist to describe sequences of dependent observations. Arguably the most popular is the ARH(1), the autoregressive Hilbertian process of order 1. It is a natural extension of the scalar and vector-valued AR(1) process (cf. Brockwell and Davis, 1991). Due to the fact that the ARH(1) model is mathematically and statistically quite flexible and well established, it is used in practice for modeling and prediction of continuous-time random experiments. We refer to Bosq (2000) for a detailed treatment of moving averages, autoregressive, and general linear time series sequences. Despite the prominent presence in time series analysis, it is clear that the applicability of moving average and autoregressive processes is limited. To describe nonlinear models in the scalar and vector cases, a number of different approaches have been introduced in the last decades. One of the most popular in econometrics is the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982) and the generalized ARCH (GARCH) model of Bollerslev (1990), which have had an enormous impact on the modeling of financial data. For surveys on volatility models we refer to Silvennoinen and Teräsvirta (2009). GARCH-type models are designed for the analysis of daily, weekly, or more general long-term period

Research by Siegfried Hörmann was partially supported by the Banque National de Belgique and Communauté française de Belgique - Actions de Recherche Concertées (2010–2015). Research by Lajos Horváth and Ron Reeder was partially supported by NSF grant DMS 0905400. Address correspondence to Siegfried Hörmann, Département de Mathématique, Université Libre de Bruxelles, CP 215, Boulevard du Triomphe, B-1050 Bruxelles, Belgium; e-mail: shormann@ulb.ac.be.

© Cambridge University Press 2012

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

returns. Improvements in data acquisition and processing techniques have led to an almost continuous flow of information for financial data with online investment decisions. High-resolution tick data are available and can be quite conveniently described as functions. It is therefore natural to ask for possible extensions of these financial time series models to a functional setup. The idea is that instead of a scalar return sequence $\{y_k, 1 \le k \le T\}$, we have a functional time series $\{y_k(t), 1 \le k \le T, 0 \le t \le S\}$, where $y_k(t)$ are intraday (log-)returns on day k at time t. In other words, if $\{P_k(t), 1 \le k \le T, 0 \le t \le S\}$ is the underlying price process, then $y_k(t) = \log P_k(t) - \log P_k(t-h)$ for the desired time lag h, where we will typically set $h = 5\,\text{min}$. By rescaling, we can always assume that $S = 1$, and then the interval [0, 1] represents one trading day.

We notice that a daily segmentation of the data is natural and preferable to only one continuous time process $\{y(s), 0 \le s \le T\}$, say, for all T days of our sample (cf. Harrison, Pitbladdo, and Schaefer, 1984; Barndorff-Nielsen and Shephard, 2004; Zhang, Mykland, and Aït-Sahalia, 2005; Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008; and Jacod, Li, Mykland, Podolskij, and Vetter, 2009). Due to the time lapse between trading days (implying, e.g., that opening and closing prices do not necessarily coincide) one continuous-time model might not be suitable for a longer period. Intraday volatilities of the euro-dollar rates investigated by Cyree, Griffiths, and Winters (2004) empirically can be considered as daily curves. Similarly, Gau (2005) studied the shape of the intraday volatility curves of the Taipei FX market. Angelidis and Degiannakis (2008) compared predictions based on intraday and interday data. Elezović (2009) modeled bid and ask prices as continuous functions. The spot exchange rates in Fatum and Pedersen (2009) can be considered as functional observations as well. Evans and Speight (2010) used 5-min returns for euro-dollar, euro-sterling and euro-yen exchange rates.

In this paper we propose a functional ARCH model. Usually time series are defined by stochastic recurrence equations establishing the relationship between past and future observations. The question preceding any further analysis is whether such an equation has a (stationary) solution. For the scalar ARCH, necessary and sufficient conditions have been derived by Nelson (1990). Interestingly, these results cannot be transferred directly to multivariate extensions (cf. Silvennoinen and Teräsvirta, 2009). Due to the complicated dynamics of multivariate ARCH/GARCH-type models (MGARCH), finding the necessary and sufficient conditions for the existence of stationary solutions to the defining equations is a difficult problem. Also, the characterization of the existence of the moments in GARCH(p, q) equations is given by very involved formulas (cf. Ling and McAleer, 2002). It is therefore not surprising that in a functional setup, i.e., when dealing with intrinsically infinite dimensional objects, some balancing between generality and mathematical feasibility of the model is required.

In Section 2 we propose a model for which we provide conditions for the existence of a unique stationary solution. These conditions are not too far from being optimal. We will also study the dependence structure of the model, which is useful

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

in many applications, e.g., in estimation, which will be treated in Section 3. We also provide an example illustrating that the proposed functional ARCH model is able to capture typical characteristics of high-frequency returns; see Section 4.

In this paper we use the following notation. Let F denote a generic function space. Throughout, this consists of real-valued functions with domain $[0,1]$. In many applications F will be equal to $\mathcal{H}=L^2([0,1])$, the Hilbert space of square integrable functions with norm $\|x\|_{\mathcal{H}} = \left(\int_0^1 x^2(s)ds\right)^{1/2}$, which is generated by the inner product $\langle x, y\rangle = \int_0^1 x(s)y(s)ds$ for $x, y \in \mathcal{H}$. Another important example is $F = \mathcal{C}[0,1]$. This is the space of continuous functions on $[0,1]$ equipped with the sup-norm $\|x\|_\infty = \sup_{t\in[0,1]} |x(t)|$. By F^+ we denote the set of nonnegative functions in F. To further lighten notation we shall often write x when we mean $\{x(t), t \in [0,1]\}$, or β for integral kernels $\{\beta(t,s), 0 \le t \le 1, 0 \le s \le 1\}$ as well as for the corresponding operators. If $x, y \in F$, then xy stands for pointwise multiplications, i.e., $xy = \{x(s)y(s), s \in [0,1]\}$. Since integrals will always be taken over the unit interval, we shall henceforth simply write $\int x(t)dt$. A random function X with values in $\mathcal{H}$ is said to be in $L^p_{\mathcal{H}}$ if $v_p(X) = \left(\mathbb{E}\|X\|^p_{\mathcal{H}}\right)^{1/p} < \infty$.

2. THE FUNCTIONAL ARCH MODEL

We start with the following general definition.

DEFINITION 2.1. *Let $\{\varepsilon_k\}$ be a sequence of independent and identically distributed random functions in F. Further let $\beta : F^+ \to F^+$ be a nonnegative operator and let $\delta \in F^+$. Then an F-valued process $\{y_k(s), k \in \mathbb{Z}, s \in [0,1]\}$ is called a functional ARCH(1) process in F if the following holds:*

$$y_k = \varepsilon_k \sigma_k \tag{2.1}$$

and

$$\sigma_k^2 = \delta + \beta(y_{k-1}^2). \tag{2.2}$$

The assumption for the existence of processes satisfying (2.1) and (2.2) depends on the choice of F. So next we specify F and put some restrictions on the operator β. Our first result gives a sufficient condition for the existence of a strictly stationary solution when $F = \mathcal{H}$. We will assume that β is a (bounded) kernel operator defined by

$$\beta(x)(t) = \int \beta(t,s)x(s)ds, \qquad x \in \mathcal{H}. \tag{2.3}$$

Boundedness is, e.g., guaranteed by finiteness of the Hilbert-Schmidt norm:

$$\|\beta\|^2_{\mathcal{S}} = \iint \beta^2(t,s)\,ds\,dt < \infty. \tag{2.4}$$

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

THEOREM 2.1. *Let $\{y_k\}$ be the process given in Definition 2.1 with $F = \mathcal{H}$ and β given in (2.3), such that the operator β is bounded. Define $K(\varepsilon_1^2) = \left(\iint \beta^2(t,s)\varepsilon_1^4(s)dsdt\right)^{1/2}$. If there is some $\alpha > 0$ such that $\mathbb{E}\{K(\varepsilon_1^2)\}^\alpha < 1$, then (2.1) and (2.2) have a unique strictly stationary solution in $\mathcal{H}$. Furthermore, σ_k^2 is of the form*

$$\sigma_k^2 = g(\varepsilon_{k-1}, \varepsilon_{k-2}, \ldots), \tag{2.5}$$

with some measurable functional $g : \mathcal{H}^{\mathbb{N}} \to \mathcal{H}$.

It follows that $\{\sigma_k\}$ and $\{y_k\}$ are not just strictly stationary but also ergodic (cf. Stout, 1974). Let $\mathcal{F}_k$ be the σ-algebra generated by the sequence $\{\varepsilon_i, i \le k\}$. If (2.1) and (2.2) have a stationary solution and if we assume that $\mathbb{E}\varepsilon_k(t) = 0$, $\mathbb{E}\varepsilon^2(t) < \infty$ and $\mathbb{E}\sigma_k^2(t) < \infty$ for all $t \in [0,1]$, then due to (2.5) it is easy to see that

$$\operatorname{corr}\{y_k(t), y_k(s)|\mathcal{F}_{k-1}\} = \operatorname{corr}\{\varepsilon_k(t), \varepsilon_k(s)\}.$$

Since by our assumption $\{\varepsilon_k, k \in \mathbb{Z}\}$ is stationary, the conditional correlation is independent of k and can be fully described by the covariance kernel $C_\varepsilon(t,s) = \operatorname{Cov}(\varepsilon(t), \varepsilon(s))$. However, we have $\operatorname{Cov}\{y_k(t), y_k(s)|\mathcal{F}_{k-1}\} = \sigma_k(t)\sigma_k(s)C_\varepsilon(s,t)$. This is in accordance with the *constant conditional correlation* (CCC) multivariate GARCH models of Bollerslev (1990) and Jeantheau (1998).

Choosing $\delta(t) = \delta$, $\beta(t,s) = \beta$, and $\varepsilon_k(t) = \varepsilon_k$ (all scalars), we get the ARCH(1) model of Nelson (1990). For this special case the condition for existence of the stationary solution according to Theorem 2.1 is that $0 < \beta < 1$ and $\mathbb{E}|\varepsilon_0|^\alpha < \infty$, which is slightly stronger than the logarithmic condition of Nelson.

Our next result shows that σ_k^2 of (2.5) can be geometrically approximated with m-dependent variables, which establishes weak dependence of the processes (2.1) and (2.2).

THEOREM 2.2. *Assume that the conditions of Theorem 2.1 hold. Let $\{\varepsilon_k'\}$ be an independent copy of $\{\varepsilon_k\}$ and define $\sigma_{km}^2 = g(\varepsilon_{k-1}, \varepsilon_{k-2}, \ldots, \varepsilon_{k-m}, \varepsilon_{k-m-1}', \varepsilon_{k-m-2}', \ldots)$. Then*

$$\mathbb{E}\{\|\sigma_k^2 - \sigma_{km}^2\|_{\mathcal{H}}\}^\alpha \le cr^m, \tag{2.6}$$

with some $0 < r = r(\alpha) < 1$ and $c = c(\alpha) < \infty$.

To better understand the idea behind our result we remark the following. Assume that we redefine

$$\sigma_{km}^2 = g\left(\varepsilon_{k-1}, \varepsilon_{k-2}, \ldots, \varepsilon_{k-m}, \varepsilon_{k-m-1,k-m}^{(k)}, \varepsilon_{k-m-2,k-m}^{(k)}, \ldots\right),$$

where $\{\varepsilon_{\ell,i}^{(k)}, \ell, i, k \in \mathbb{Z}\}$ are independent copies of $\{\varepsilon_\ell, \ell \in \mathbb{Z}\}$. In other words, every σ_k^2 gets its "individual" copy of $\{\varepsilon_{\ell,i}^{(k)}\}$ to define the approximations. It can be easily seen that then for any fixed $m \ge 1$, $\{\sigma_{km}^2, k \in \mathbb{Z}\}$ form m-dependent sequences,

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

while the value on the left-hand side in inequality (2.6) doesn't change. As we have shown in recent papers (Aue, Hörmann, Horváth, Hůsková, and Steinebach, 2012; Hörmann and Kokoszka, 2010), approximations like (2.6) are particularly useful in studying large sample properties of functional data. We use (2.6) to provide conditions for the existence of moments of the stationary solutions of (2.1) and (2.2). It also follows immediately from (2.6), that if (2.1) and (2.2) are solved starting with some initial values y_0^* and σ_0^*, then the effect of the initial values dies out exponentially fast.

In a finite-dimensional vector space, all norms are equivalent. This is no longer true in the functional (infinite-dimensional) setup, and whether a solution of (2.1) and (2.2) exists depends on the choice of space and norm of the state space. Depending on the application, it might be more convenient to work in a different space. We give here the analogue of Theorems 2.1 and 2.2 for a functional ARCH process in $\mathcal{C}[0,1]$.

THEOREM 2.3. *Let $\{y_k\}$ be the process given in Definition 2.1 with $F = \mathcal{C}[0,1]$ and define $H(\varepsilon_1^2) = \sup_{0\le t\le 1}\int \beta(t,s)\varepsilon_1^2(s)ds$. If there is some $\alpha > 0$ such that $\mathbb{E}\{H(\varepsilon_1^2)\}^\alpha < 1$, then (2.1) and (2.2) have a unique strictly stationary solution in $\mathcal{C}[0,1]$. Furthermore, σ_k^2 can be represented as in (2.5). In addition, the proposition of Theorem 2.2 holds, with (2.6) replaced by*

$$\mathbb{E}\{\|\sigma_k^2 - \sigma_{km}^2\|_\infty\}^\alpha \le cr^m.$$

We continue with some immediate consequences of our theorems. We start with conditions for the existence of the moments of the stationary solution of (2.1) and (2.2).

PROPOSITION 2.1. *Assume that the conditions of Theorem 2.1 hold. Then*

$$\mathbb{E}\{\|\sigma_0^2\|_{\mathcal{H}}\}^\alpha < \infty. \tag{2.7}$$

If

$$\mathbb{E}\{\|\sigma_0\|_{\mathcal{H}}\}^\alpha < \infty \tag{2.8}$$

and

$$\mathbb{E}\{\|\varepsilon_0\|_\infty\}^\alpha < \infty, \tag{2.9}$$

then

$$\mathbb{E}\{\|y_0\|_{\mathcal{H}}\}^\alpha < \infty. \tag{2.10}$$

PROPOSITION 2.2. *Assume that the conditions of Theorem 2.3 hold. Then the analogue of Proposition 2.1 holds, with $\|\cdot\|_{\mathcal{H}}$ in (2.7)–(2.10) replaced by $\|\cdot\|_\infty$.*

We would like to point out that it is not assumed that the innovations ε_k have finite variance. We only need that ε_k have some moment of order $\alpha > 0$, where

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

$\alpha > 0$ can be as small as we wish. Hence our model allows for innovations as well as observations with heavy tails.

According to Propositions 2.1 and 2.2, if the innovation ε_0 has enough moments, then so do σ_0^2 and y_0. The next result shows a connection between the moduli of continuity of ε_0 and y_0. Let

$$\omega(x,h) = \sup_{0\le t\le 1-h} \sup_{0\le s\le h} |x(t+s)-x(t)|$$

denote the modulus of continuity of a function $x(t)$.

PROPOSITION 2.3. *We assume that the conditions of Theorem 2.3 are satisfied with* $\alpha = p > 0$. *If* $\mathbb{E}\{\|\varepsilon_0\|_\infty\}^p < \infty$ *and* $\lim_{h\to 0}\mathbb{E}\{\omega(\varepsilon_0,h)\}^p = 0$, *then* $\lim_{h\to 0}\mathbb{E}\{\omega(y_0,h)\}^p = 0$.

According to Theorems 2.2 and 2.3, the stationary solution of (2.1) and (2.2) can be approximated with stationary, weakly dependent sequences with values in $\mathcal{H}$ and in $\mathcal{C}[0,1]$, respectively. We provide two further results that establish the weak dependence structure of $\{y_k\}$.

PROPOSITION 2.4. *We assume that the conditions of Theorem 2.1 are satisfied with* $\alpha = p/2$ *and*

$$\mathbb{E}\{\|\varepsilon_0\|_\infty\}^p < \infty. \tag{2.11}$$

Then

$$\mathbb{E}\{\|y_k - y_{km}\|_{\mathcal{H}}\}^p \le c\gamma^m, \qquad -\infty < k < \infty,\ m \ge 1, \tag{2.12}$$

with some $0 < c < \infty$ *and* $0 < \gamma < 1$, *where* $y_{km} = \varepsilon_k\sigma_{km}$.

It follows from the definitions that the distribution of the $y_k - y_{km}$ does not depend on k. Hence the expected value in (2.12) does not depend on k. A similar result holds in $F = \mathcal{C}[0,1]$ under the sup-norm.

PROPOSITION 2.5. *We assume that the conditions of Theorem 2.3 are satisfied with* $\alpha = p/2$ *and that (2.11) holds. Then*

$$\sup_{0\le t\le 1} \mathbb{E}|y_k - y_{km}|^p \le c\gamma^m, \qquad -\infty < k < \infty,\ m \ge 1, \tag{2.13}$$

with some $0 < c < \infty$ *and* $0 < \gamma < 1$, *where* $y_{km} = \varepsilon_k\sigma_{km}$.

As in case of Proposition 2.4, the expected value in (2.13) does not depend on k.

3. ESTIMATION

In this section we propose estimators for the function δ and the operator β in models (2.1)–(2.2), which are not known in practice. These estimators are based on

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

N observations. The procedure is developed for the important case where $F = \mathcal{H}$ and β is given as in (2.3). We show that our problem is related to the estimation of the autocorrelation operator in the ARH(1) model, which has been intensively studied in Bosq (2000). However, the theory developed in Bosq is not directly applicable, as it requires independent innovations in the ARH(1) process, whereas, as we will see below, we can only assume weak white noise (in Hilbert space sense).

We will impose the following.

Assumption 3.1.

(a) $\mathbb{E}\varepsilon_0^2(t) = 1$ for any $t \in [0, 1]$.
(b) The assumptions of Theorem 2.2 hold with $\alpha = 2$.

Assumption 3.1(a) is needed to guarantee the identifiability of the model. Part (b) of the assumption guarantees the existence of a stationary solution of the model (2.1)–(2.2) with moments of order 4. It is necessary to make the moment-based estimator proposed below working. An immediate consequence of Assumption 3.1 is that (2.4) holds; i.e., β is a Hilbert Schmidt operator.

We let m_2 denote the mean function of the y_k^2 and introduce

$$\nu_k = y_k^2 - \sigma_k^2 = \{(\varepsilon_k^2(s) - 1)\sigma_k^2(s),\, s \in [0, 1]\}.$$

Then by adding ν_k on both sides of (2.2) we obtain

$$y_k^2 = \delta + \beta(y_{k-1}^2) + \nu_k.$$

Since β is a linear operator, we obtain after subtracting m_2 on both sides of the above equation,

$$y_k^2 - m_2 = \delta - m_2 + \beta(m_2) + \beta(y_{k-1}^2 - m_2) + \nu_k. \tag{3.1}$$

It can be easily seen that under Assumption 3.1, $\mathbb{E}\nu_k = 0$ (where 0 stands for the zero function). Notice also that the expectation commutes with bounded operators, and hence that $\mathbb{E}(\beta(y_k^2 - m_2)) = \beta(\mathbb{E}(y_k^2 - m_2)) = 0$. Consequently, taking expectations on both sides of (3.1) yields that

$$\delta - m_2 + \beta(m_2) = 0. \tag{3.2}$$

Thus, (3.1) can be rewritten in the form

$$Z_k = \beta(Z_{k-1}) + \nu_k \quad \text{with} \quad Z_k = y_k^2 - m_2. \tag{3.3}$$

Model (3.3) is the autoregressive Hilbertian model of order 1, short ARH(1). For estimating the autocorrelation operator β we may use the estimator proposed in Bosq (2000, Chap. 8). We need to be aware, however, that the theory in Bosq has been developed for ARH processes with strong white noise innovations,

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

i.e., independent innovations $\{v_k\}$. In our setup the $\{v_k\}$ form only a *weak white noise* sequence; i.e., for any $n \neq m$ we have

$$\mathbb{E}\|v_n\|_{\mathcal{H}}^2 < \infty \quad \text{and} \quad \mathbb{E}\langle v_n, x\rangle\langle v_m, y\rangle = 0 \quad \forall x, y \in \mathcal{H},$$

and the covariance operator of v_n is independent of n. Thus the theory in Bosq cannot be directly applied. We will study the estimation of β in Section 3.1.

Once β is estimated by some $\hat{\beta}$, say, we obtain an estimator for δ via equation (3.2):

$$\hat{\delta} = \hat{m}_2 - \hat{\beta}(\hat{m}_2), \tag{3.4}$$

where we use

$$\hat{m}_2 = \frac{1}{N}\sum_{k=1}^{N} y_k^2. \tag{3.5}$$

Let $\|\beta\|_{\mathcal{L}} = \sup_{x\in\mathcal{H}}\{\|\beta(x)\|_{\mathcal{H}} : \|x\| \leq 1\}$ be the operator norm of β. Recall that $\|\beta\|_{\mathcal{L}} \leq \|\beta\|_{\mathcal{S}}$. The following lemma shows that consistency of $\hat{\beta}$ implies consistency of $\hat{\delta}$.

LEMMA 3.1. *Let Assumption 3.1 hold. Then*

$$\|\hat{\delta} - \delta\|_{\mathcal{H}} = O_P(1) \times \left(N^{-1/2} + \|\hat{\beta} - \beta\|_{\mathcal{L}}\right).$$

Proof. We have

$$\begin{aligned}\|\hat{\delta} - \delta\|_{\mathcal{H}} &\leq \|\hat{m}_2 - m_2\|_{\mathcal{H}} + \|\hat{\beta}(\hat{m}_2) - \beta(\hat{m}_2)\|_{\mathcal{H}} + \|\beta(\hat{m}_2) - \beta(m_2)\|_{\mathcal{H}} \\ &\leq \|\hat{m}_2 - m_2\|_{\mathcal{H}} + \|\hat{\beta} - \beta\|_{\mathcal{L}}\|\hat{m}_2\|_{\mathcal{H}} + \|\beta\|_{\mathcal{L}}\|\hat{m}_2 - m_2\|_{\mathcal{H}}.\end{aligned}$$

The result follows once we can show that $\|\hat{m}_2 - m_2\|_{\mathcal{H}} = \|\hat{m}_{2,N} - m_2\|_{\mathcal{H}} = O_P(N^{-1/2})$. To this end we notice that by stationarity of $\{y_k^2\}$,

$$\begin{aligned}\mathbb{E}\|\hat{m}_2 - m_2\|_{\mathcal{H}}^2 &= \mathbb{E}\left\|\frac{1}{N}\sum_{k=1}^{N}(y_k^2 - m_2)\right\|_{\mathcal{H}}^2 \\ &= \frac{1}{N}\sum_{k=-(N-1)}^{N-1}\left(1 - \frac{|k|}{N}\right)\mathbb{E}\left\langle y_0^2 - m_2, y_k^2 - m_2\right\rangle \\ &\leq \frac{1}{N}\left(\mathbb{E}\left\|y_0^2 - m_2\right\|_{\mathcal{H}} + 2\sum_{k=1}^{\infty}\left|\mathbb{E}\left\langle y_0^2 - m_2, y_k^2 - m_2\right\rangle\right|\right).\end{aligned}$$

By construction, y_0^2 and the approximation y_{kk}^2 are independent. Repeated application of the Cauchy-Schwarz inequality together with Assumption 3.1(a) yield that

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

$$\begin{aligned}\left|\mathbb{E}\left\langle y_0^2 - m_2, y_k^2 - m_2\right\rangle\right| &= \left|\mathbb{E}\left\langle y_0^2 - m_2, y_k^2 - y_{kk}^2\right\rangle\right| \\ &\le \left(\mathbb{E}\left\|y_0^2 - m_2\right\|_{\mathcal{H}}^2\right)^{1/2}\left(\mathbb{E}\left\|y_k^2 - y_{kk}^2\right\|_{\mathcal{H}}^2\right)^{1/2} \\ &= \left(\mathbb{E}\left\|\sigma_0^2 - m_2\right\|_{\mathcal{H}}^2\right)^{1/2}\left(\mathbb{E}\left\|\sigma_k^2 - \sigma_{kk}^2\right\|_{\mathcal{H}}^2\right)^{1/2}.\end{aligned}$$

Combining these estimates with Theorem 2.2 shows that $\mathbb{E}\left\|\hat{m}_2 - m_2\right\|_{\mathcal{H}} = O(N^{-1/2})$. ■

3.1. Estimation of β

We now turn to the estimation of the autoregressive operator β in the ARH(1) model (3.3). It is instructive to focus first on the univariate case $Z_n = \beta Z_{n-1} + \nu_n$, in which all quantities are scalars. We assume $\mathbb{E}\nu_n = 0$, which implies $\mathbb{E}Z_n = 0$. We also assume that $|\beta| < 1$, so that there is a stationary solution such that ν_n is uncorrelated with Z_{n-1}. Then, multiplying the AR(1) equation by Z_{n-1} and taking the expectation, we obtain $\gamma_1 = \beta\gamma_0$, where $\gamma_k = \mathbb{E}[Z_n Z_{n+k}] = \operatorname{cov}(Z_n, Z_{n+k})$. The autocovariances γ_k are estimated in the usual way by the sample autocovariances

$$\hat{\gamma}_k = \frac{1}{N}\sum_{j=1}^{N-k} Z_j Z_{j+k},$$

so the usual estimator of β is $\hat{\beta} = \hat{\gamma}_1/\hat{\gamma}_0$. This is the so-called *Yule-Walker estimator*, which is optimal in many ways (see Brockwell and Davis, 1991, Chap. 8).

In the functional setup we will replace condition $|\beta| < 1$ with $\|\beta\|_{\mathcal{S}} < 1$. Notice that this condition is guaranteed by Assumption 3.1 and that it will imply the existence of a weakly stationary solution of (3.3) of the form

$$Z_n = \sum_{j\ge 0} \beta^j(\nu_{n-j}),$$

where β^j is the j-times iteration of the operator β and β^0 is the identity mapping. The estimator for the operator β obtained in Bosq (2000) is formally analogue to the scalar case. We need instead of γ_0 and γ_1 the covariance operator

$$C_0(\cdot) = \mathbb{E}[\langle Z_1, \cdot\rangle Z_1]$$

and the cross-covariance operator

$$C_1(\cdot) = \mathbb{E}[\langle Z_1, \cdot\rangle Z_2].$$

One can show by similar arguments as in the scalar case that

$$\beta = C_1 C^{-1}.$$

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

To get an explicit form, let $\lambda_1 \geq \lambda_2 \geq \cdots$ be the eigenvalues of C and let $e_1, e_2, \ldots$ be the corresponding eigenfunctions; i.e., $C(e_i) = \lambda_i e_i$. We assume that e_j are normalized to satisfy $\|e_j\|_{\mathcal{H}} = 1$. Then $\{e_j\}$ forms an orthonormal basis (ONB) of $\mathcal{H}$, and we obtain the spectral decomposition of the operator C,

$$C(y) = \sum_{j \geq 1} \lambda_j \langle e_j, y \rangle e_j. \tag{3.6}$$

From (3.6) we get formally that

$$C^{-1}(y) = \sum_{j \geq 1} \lambda_j^{-1} \langle e_j, y \rangle e_j, \tag{3.7}$$

and hence

$$\begin{aligned} \beta(y) = C_1 C^{-1}(y) &= \mathbb{E}\left(\left\langle Z_1, \sum_{j \geq 1} \lambda_j^{-1} \langle e_j, y \rangle e_j \right\rangle Z_2 \right) \\ &= \sum_{j \geq 1} \lambda_j^{-1} \langle e_j, y \rangle \mathbb{E}\left(\langle Z_1, e_j \rangle Z_2 \right). \end{aligned} \tag{3.8}$$

Using $Z_2 = \sum_{i \geq 1} \langle Z_2, e_i \rangle e_i$, we obtain that the corresponding kernel is

$$\beta(t,s) = \sum_{i,j \geq 1} \lambda_j^{-1} \mathbb{E}\left(\langle Z_1, e_j \rangle \langle Z_2, e_i \rangle \right) e_j(s) e_i(t). \tag{3.9}$$

If $\lambda_j = 0$ for all $j > p \geq 1$, then the covariance operator is finite rank and we can replace (3.6) and (3.7) by finite expansions with the sum going from 1 to p. In this case, all our mathematical operations so far are well justified. However, when all $\lambda_j > 0$, then we need to be aware that C^{-1} is not bounded on $\mathcal{H}$. To see this, note that $\lambda_j \to 0$ if $j \to \infty$ (this follows from the fact that C is a Hilbert-Schmidt operator). Consequently, $\|C^{-1}(e_j)\|_{\mathcal{H}} = \lambda_j^{-1} \to \infty$ for $j \to \infty$. It can be easily seen that this operator is bounded only on

$$D = \left\{ y \in \mathcal{H} : \sum_{j \geq 1} \frac{\langle e_j, y \rangle^2}{\lambda_j^2} < \infty \right\}.$$

Nevertheless, we can show that the representation (3.8) holds for all $y \in \mathcal{H}$ by using a direct expansion of $\beta(t,s)$. Since the eigenfunctions $\{e_k, k \geq 1\}$ of C form an ONB of $\mathcal{H}$, it follows that $\{e_k(s)e_\ell(t), k, \ell \geq 1\}$ forms an ONB of $L^2([0,1]^2) = \mathcal{H}^2$. This is again a Hilbert space with inner product

$$\langle x, y \rangle_{\mathcal{H}^2} = \int \int x(t,s) y(t,s) dt ds.$$

Note that $\|\beta\|_{\mathcal{H}^2} = \|\beta\|_{\mathcal{S}} < \infty$ and hence the kernel function $\beta \in \mathcal{H}^2$. (Be aware that for the sake of a lighter notation we don't distinguish between kernel and operator β.) Consequently, $\beta(t,s)$ has the representation

$$\beta(t,s) = \sum_{k,\ell \geq 1} \beta_{k,\ell} e_k(t) e_\ell(s). \tag{3.10}$$

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

As we can write

$$Z_{n+1} = \sum_{k,\ell \geq 1} \beta_{k,\ell} \langle Z_n, e_k \rangle e_\ell + \nu_{n+1},$$

it follows that

$$\langle Z_{n+1}, e_i \rangle \langle Z_n, e_j \rangle = \sum_{k \geq 1} \beta_{k,i} \langle Z_n, e_k \rangle \langle Z_n, e_j \rangle + \langle \nu_{n+1}, e_i \rangle \langle Z_n, e_j \rangle,$$

and, by taking expectations on both sides of the above equation, that

$$\mathbb{E}\langle Z_2, e_i \rangle \langle Z_1, e_j \rangle = \sum_{k \geq 1} \beta_{k,i} \langle C(e_k), e_j \rangle = \beta_{j,i} \lambda_j. \tag{3.11}$$

Here we used the fact that $\{\nu_k\}$ is weak white noise. It implies that $\mathbb{E}\langle B(\nu_k), x \rangle \langle \nu_\ell, y \rangle$ is zero for any bounded operator B and all $x, y \in \mathcal{H}$ and all $k \neq \ell$. Hence the expansion $Z_k = \sum_{j \geq 0} \beta^j (\nu_{k-j})$ provides $\mathbb{E}\langle \nu_{n+1}, e_i \rangle \langle Z_n, e_j \rangle = 0$. Hence (3.9) follows from (3.10) and (3.11).

Now we would like to obtain an estimator for β by using a finite sample version of the above relations. To this end we set

$$\hat{C}(y) = \frac{1}{N} \sum_{k=1}^{N} \langle Z_k, y \rangle Z_k \quad \text{and} \quad \hat{C}_1(y) = \frac{1}{N} \sum_{k=1}^{N-1} \langle Z_k, y \rangle Z_{k+1}, \qquad y \in \mathcal{H}.$$

The estimator in Bosq (2000) and the estimator we also propose here is of the form

$$\hat{\beta}(y; K) = \pi_K \hat{C}_1 \widehat{C^{-1}}(y; K),$$

where

$$\widehat{C^{-1}}(y; K) = \sum_{j=1}^{K} \hat{\lambda}_j^{-1} \langle \hat{e}_j, y \rangle \hat{e}_j, \tag{3.12}$$

$(\hat{\lambda}_j, \hat{e}_j)$ are the eigenvalues (in descending order) and the corresponding eigenfunctions of $\hat{C}$, and p_K is the orthogonal projection onto the subspace $\text{span}(\hat{e}_1, \ldots, \hat{e}_K)$. We notice that this estimator is not depending on the sign of the $\hat{e}_j$'s. The corresponding kernel is given as

$$\hat{\beta}(t, s; K) = \frac{1}{N-1} \sum_{k=1}^{N-1} \sum_{j=1}^{K} \sum_{i=1}^{K} \hat{\lambda}_j^{-1} \langle Z_k, \hat{e}_j \rangle \langle Z_{k+1}, \hat{e}_i \rangle \hat{e}_j(s) \hat{e}_i(t), \tag{3.13}$$

and the signs of the $\hat{e}_j$ cancel out. In practice, eigenvalues and eigenfunctions of an empirical covariance operator can be conveniently computed with the package `fda` for the statistical software `R`. The estimator (3.13) is the empirical version of the finite expansion

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

$$\beta(t,s;K)=\sum_{i=1}^{K}\sum_{j=1}^{K}\lambda_j^{-1}\mathbb{E}\left(\langle Z_1,e_j\rangle\langle Z_2,e_i\rangle\right)e_j(s)e_i(t)$$

of (3.9).

If the innovations $\{\nu_k\}$ are i.i.d., Bosq (2000) proves under some technical conditions consistency of the estimator (3.13) when $K=K(N)$:

$$\|\beta-\hat{\beta}(K(N))\|_{\mathcal{L}}=o_P(1)\quad\text{as } N\to\infty. \tag{3.14}$$

The result in (3.14) holds for a functional ARCH process as well. However, the choice of $K(N)$ depends on the decay rate of the eigenvalues, which is unknown in practice and is determined by the unknown values of $\delta(t)$ and $\beta(t,s)$. Empirical results (see Didericksen, Kokoszka, and Zhang, 2010) show that in the finite sample case, $K=2,3,4$ provides best results. Choosing small K is often favorable due to a bias variance trade-off. Note that the eigenvalues occur reciprocal in the estimator $\hat{\beta}$, and thus larger K accounts for larger instability if the eigenvalues are close to zero. A practical approach is to choose K the largest integer for which $\hat{\lambda}_K/\hat{\lambda}_1\geq\gamma$, where γ is some threshold.

THEOREM 3.1. *Fix some $K\geq 1$. Assume that the $K+1$ largest eigenvalues of the covariance operator C of Z_k satisfy $\lambda_1>\lambda_2>\cdots>\lambda_{K+1}>0$. Let $\beta(K)$ and $\hat{\beta}(K)$ be the operators belonging to the kernel functions $\beta(t,s;K)$ and $\hat{\beta}(t,s;K)$, respectively. Let Assumption 3.1 hold with condition (b) strengthened to $\alpha=4$. Then we have*

$$\|\beta(K)-\hat{\beta}(K)\|_{\mathcal{S}}=O_P\left(N^{-1/2}\right)\quad \textit{as } N\to\infty.$$

In Theorem (3.1) N obviously denotes the sample size that is suppressed in the notation. The proof of this theorem is given in Section 5. Our conditions imply that $\mathbb{E}\|Z_k\|^4<\infty$. This assumption is probably more stringent than necessary, and a relaxation would be desirable. Note, however, that finite 4th moments are required in Bosq (2000) even for i.i.d. $\{\nu_k\}$.

Yao and Müller (2010) suggested cross-validation to choose K. Let $\hat{\beta}_{(-i)}(t,s,K)$ be the estimator for $\beta(t,s)$ obtained by leaving the ith observation out of the sample. The analogous estimator for $\delta(t)$ will be $\hat{\delta}_{(-i)}(t,K)$. We define $\hat{\sigma}^2_{i,K}(t)=\hat{\delta}_{(-i)}(t,K)+\int\hat{\beta}_{(-i)}(t,s,K)y^2_{i-1}(s)ds$. Since the volatility function, σ_i^2, is not observed, we use the observed volatility, y_i^2, in its place. A cross-validation-based criteria for choosing K is to minimize

$$A_N(K)=\frac{1}{N}\sum_{i=1}^{N}\int\left|y_i^2(t)-\hat{\sigma}^2_{i,K}(t)\right|^{\kappa}dt, \tag{3.15}$$

where $\kappa\geq 1$ and usually chosen to be 1. Very little is known about the asymptotic behavior of cross-validation even when the functional observations are independent and identically distributed. It would be an interesting but very

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

complex project to get statistical inference on the cross-validated $K(N)$ in case of dependent functional observations as in our model.

The sum of the one-leave-out prediction errors, $A_N(K)$ in our case, has also been suggested to evaluate the fit of a model to the data in case of a simple situation by Yao and Müller (2010). An analogue of $A_N(K)$ in case of a functional autoregressive model was investigated numerically by Didericksen et al. (2010). They observed that the parameter estimation-based method worked as well or better than the more complex method developed by Kargin and Onatski (2008) for prediction in autoregressive functional models. We think that the Kargin–Onatski method can also be adapted to our function ARCH model.

3.2. Simulation Study

In this section we demonstrate the capabilities of our estimators for $\beta(t,s)$ and $\delta(t)$ on simulated data. We proceed as follows: We choose a simple $\beta(t,s)$ and $\delta(t)$, simulate several days of observations using these parameters, and then use the estimation procedure given in Section 3.1 to obtain $\hat{\beta}(t,s;2)$ and $\hat{\delta}(t;2)$ from (3.13) and (3.4), respectively. Then we will compare the true values of the parameters to our estimates.

We will use $\beta(t,s) = 16s(1-s)t(1-t)$ and $\delta(t) = 0.01$ for our simulations. Now that we have chosen $\beta(t,s)$, and $\delta(t)$, we can simulate data according to (2.1) and (2.2). We will use $\varepsilon_i(t) = B_i(t) + N_i\sqrt{1-t(1-t)}$ for the error term, where $B_i(t)$ are i.i.d. standard Brownian bridges and N_i are i.i.d. standard normals. Note that this gives $\mathbb{E}(\varepsilon^2(t)) = 1$ for all t, which is assumed by our estimation procedure. After simulating N days of data, we compute $\hat{\beta}(t,s;2)$ and $\hat{\delta}(t;2)$. A plot of the true $\beta(t,s)$ is given in Figure 1. Figure 2 shows the estimates when $N = 300$.

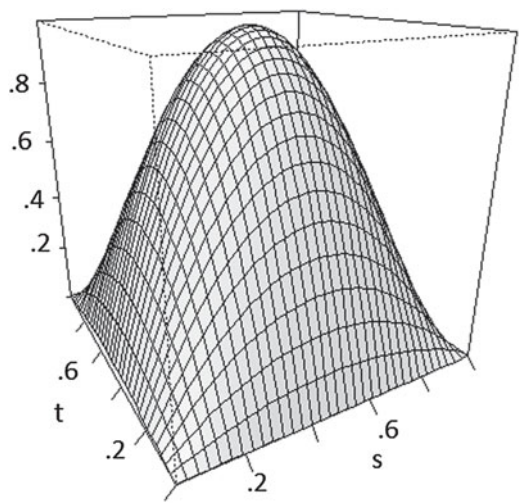


FIGURE 1. $\beta(t,s) = 16s(1-s)t(1-t)$.

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

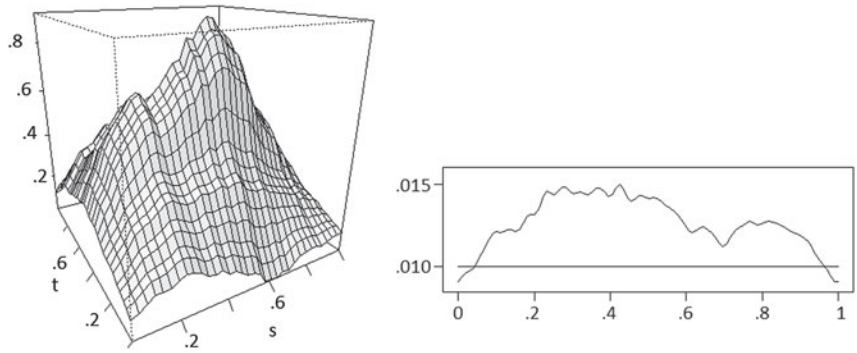


FIGURE 2. Using a sample of size $N = 300$, we obtain $\hat{\beta}(t, s; 2)$ on the left and $\hat{\delta}(t; 2)$ with $\delta(t) = .01$ superimposed on the right.

We see from these plots that the estimators described in Section 3.1 accurately estimate the parameters, $\beta(t, s)$ and $\delta(t)$, when the sample size is sufficiently large. Note that the plot of $\hat{\delta}(t; 2)$ has the true $\delta(t)$ superimposed.

4. AN EXAMPLE

In this section we show an example illustrating that our model captures some basic features of intraday returns. Let $P_k(t)$ denote the price of a stock on day k at time t. Then $y_k(t)$ can be viewed as the log-returns of the stock, $y_k(t) = \log P_k(t) - \log P_k(t-h)$, during period h (cf. Cyree et al., 2004), where h is typically 1, 5, or 15 minutes. We will use $h = 5$ for 5-minute returns. For the S&P 100 index, the 5-minute log-return process, $y_k(t)$, is recorded at 1-minute intervals throughout each trading day. The volatility of the stock is then represented by $\sigma_k^2(t) = \mathrm{Var}(y_k(t)|\mathcal{F}_{k-1})$. If we were to use a multivariate ARCH or GARCH model, it would have a dimension of 400, making it extremely impractical to model the volatility of the returns.

The first step to simulating the intraday returns is to estimate the parameters, $\delta(t)$ and $\beta(t, s)$, as outlined in Section 3.1. The first 400 observations were used to compute $A_{400}(K)$ of (3.15) with $\kappa = 2$. We obtained 1.1760×10^{-6} $(p = 1)$, $1.1709 \times 10^{-6} (p = 2)$, and $1.1709 \times 10^{-6} (p = 3)$, so A_{400} has stabilized at $p = 2$. The estimated functions, $\hat{\beta}(t, s; 2)$ and $\hat{\delta}(t; 2)$, are shown in Figure 3.

Notice in Figure 3 that $\hat{\beta}(t, s; 2)$ and $\hat{\delta}(t; 2)$ are somewhat larger when t is close to 0 or 1. According to (2.2), this suggests that the volatility, $\sigma_k^2(t)$, tends to be larger at the beginning and end of each trading day. Higher volatilities at the beginning and the end of the trading day have been observed by several authors (cf. Gau, 2005; and Evans and Speight, 2010). This phenomenon is consistent with our observed log-return data based on the S&P 100 index and is captured by our model.

Having estimated the parameters, $\delta(t)$ and $\beta(t, s)$, we can now simulate several days of observations according to (2.1) and (2.2). We will use

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

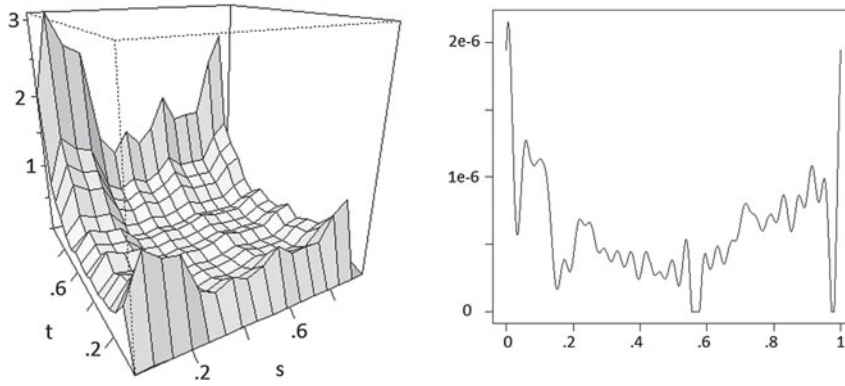


FIGURE 3. Left: $\hat{\beta}(t,s;2)$ estimated from S&P 100 index. Right: $\hat{\delta}(t;2)$ estimated from S&P 100 index.

$\varepsilon_i(t) = 2^{-200t}\sqrt{\log(2)}W_i(2^{400t}/\log(2))$ for the error term, where $W_i(t)$ are i.i.d. standard Brownian motions. Note that this gives $\mathbb{E}(\varepsilon^2(t)) = 1$ for all t, which is assumed by our estimation procedure.

We simulated two days of log-returns, which we compare with the log-returns of the S&P 100 index. The right side of Figure 4 is the plot of the 5-minute returns on the S&P 100 index on April 11 and 12, 2000. The left side of Figure 4 shows two consecutive days of simulated values for $y_k(t)$. The simulations show that our model empirically captures some of the characteristics of financial data. In particular, there are higher volatilities at the beginning and end of the trading day.

We also used our model to predict volatility in the S&P 100 data. After having observed $y_1 \cdots, y_n$, we predict the volatility at time $n+1$ with

$$\hat{\sigma}_n^2(t,1) = \hat{\delta}_n(t) + \int_0^1 \hat{\beta}_n(t,s)y_n^2(s)ds$$

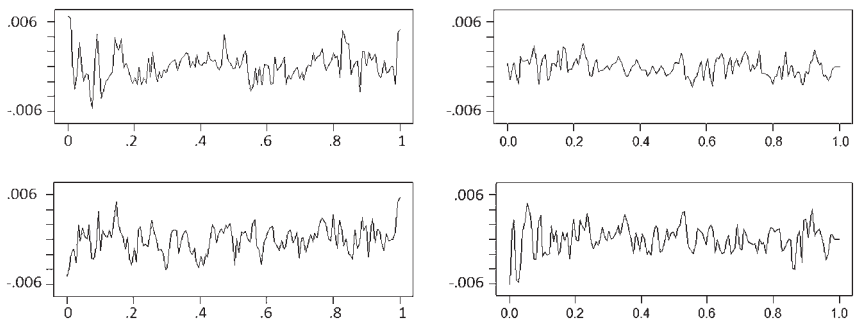


FIGURE 4. Left panel: Two consecutive days of simulated values for $y_k(t)$. Right panel: 5-minute log-returns for the S&P index on April 11 and 12, 2000.

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

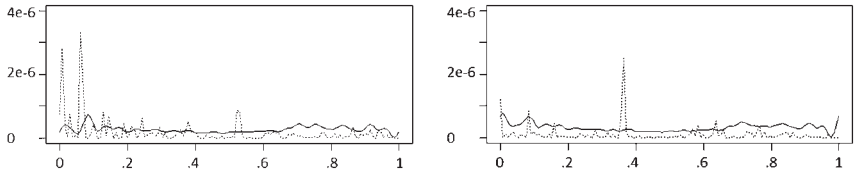


FIGURE 5. $\hat{\sigma}^2_{500}(t,1)$ (solid line) is superimposed onto the plot of the associated $y^2_{501}(t)$ (dashed line) for April 3, 2007 (left panel), and $\hat{\sigma}^2_{500}(t,2)$ (solid line) is superimposed onto the plot of the associated $y^2_{502}(t)$ (dashed line) for April 4, 2007 (right panel).

and with

$$\hat{\sigma}^2_n(t,2) = \hat{\delta}_n(t) + \int_0^1 \hat{\beta}_n(t,s)\hat{\sigma}^2_n(s,1)ds$$

for time $n+2$, where $\hat{\delta}_n$ and $\hat{\beta}_n$ are the corresponding estimators of the parameters. Using the observations collected for the period of April 12, 2005, to April 2, 2007, we predicted the volatility curves for April 3 and 4, 2007. The predictors and the observed volatilities are exhibited in Figure 5. For a more detailed simulation study, we refer to Reeder (2011).

5. PROOFS

The proofs of Theorems 2.1 and 2.3 are based on general results for iterated random functions as those in Wu and Shao (2004) and Diaconis and Freedman (1999). For the convenience of the reader, we shall repeat here the main ideas of Wu and Shao.

Let (S,ρ) be a complete, separable metric space. Let Θ be another metric space and let $M : \Theta \times S \to S$ be a measurable function. For a random element θ with values in Θ, an iterated random function system is defined via the random mappings $M_\theta(\cdot)$. More precisely, it is assumed that

$$X_n = M_{\theta_n}(X_{n-1}), \qquad n \in \mathbb{N}, \tag{5.1}$$

where $\{\theta_n,\ -\infty < n < \infty\}$ is an i.i.d. sequence with values in Θ. Thereby it is assumed that X_0 is independent of $\{\theta_n, n \geq 1\}$. For any $x \in S$, we define

$$S_n(x) = M_{\theta_n} \circ M_{\theta_{n-1}} \circ \cdots \circ M_{\theta_1}(x), \qquad n = 1,2,\ldots,$$

where $\circ$ denotes the composition of functions. We also introduce the backward version of S_n, which is given by

$$Z_n(x) = M_{\theta_{-1}} \circ M_{\theta_{-2}} \circ \cdots \circ M_{\theta_{-n}}(x), \qquad x \in S,\ n = 1,2,\ldots.$$

The following theorem is a slight modification of Theorem 2 of Wu and Shao (2004), so it is immediately applicable for our purposes.

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

THEOREM 5.1 *(Wu and Shao, 2004). Assume that*

(i) there are $y_0 \in S$ *and* $\alpha > 0$ *such that* $\mathbb{E}\{\rho(y_0, M_{\theta_0}(y_0))\}^\alpha < \infty$ *and*
(ii) there are $x_0 \in S$, $\alpha > 0$, $0 < r_1 = r_1(\alpha) < 1$ *and* $c = c(\alpha) < \infty$ *such that*

$$\mathbb{E}\{\rho(S_n(x), S_n(x_0))\}^\alpha \le c r_1^n \{\rho(x, x_0)\}^\alpha$$

for all $x \in S$ *and* $n \in \mathbb{N}$. *Then for all* $x \in S$, *we have* $Z_n(x)$ *converges almost surely to some* Z_∞, *which is independent of* x. *Furthermore,* $Z_\infty = g(\theta_0, \theta_{-1}, \ldots)$ *and*

$$\mathbb{E}\{\rho(Z_n(x), Z_\infty)\}^\alpha \le c_1 r^n,$$

where $c_1 = c_1(x, x_0, y_0, \alpha) < \infty$ *and* $0 < r = r(\alpha) < 1$. *Moreover, the process* $X_n = g(\theta_n, \theta_{n-1}, \ldots)$ *is a stationary solution of (5.1). Finally, if we let* $X_0^* = f(\theta_0', \theta_{-1}', \ldots)$, *where* $\{\theta_n'\}$ *is an independent copy of* $\{\theta_n\}$, *then*

$$\mathbb{E}\{\rho(S_n(X_0^*), S_n(X_0))\}^\alpha \le c_2 r_2^n,$$

with some $0 < r_2 = r_2(\alpha) < 1$ *and* $c_2 = c_2(\alpha) > 0$.

Proof of Theorem 2.1. We need to show that the conditions of Theorem 5.1 are satisfied when the underlying space is $\mathcal{H}$ with metric $\|\cdot\|_{\mathcal{H}}$ and

$$M_{\theta_n}(x)(t) = \delta(t) + \int \beta(t,s)\varepsilon_{n-1}^2(s)x(s)ds.$$

To demonstrate Theorem 5.1(i), we use $y_0(t) = 0$, $0 \le t \le 1$, and get

$$\|y_0 - M_{\theta_0^2}(y_0)\|_{\mathcal{H}}^2 = \int \delta^2(t)dt < \infty,$$

by assumption. Since for any $x, x_0 \in \mathcal{H}$ we have

$$\begin{aligned}
&\|S_n(x) - S_n(x_0)\|_{\mathcal{H}} \\
&\quad = \|M_{\theta_n^2}(S_{n-1}(x)) - M_{\theta_n^2}(S_{n-1}(x_0))\|_{\mathcal{H}} \\
&\quad = \left(\int \left(\int \beta(t,s)\{S_{n-1}(x)(s) - S_{n-1}(x_0)(s)\}\varepsilon_{n-1}^2(s)ds \right)^2 dt \right)^{1/2} \\
&\quad \le \left(\int \left\{ \int \beta^2(t,s)\varepsilon_{n-1}^4(s)ds \right\} \int \left\{ S_{n-1}(x)(s) - S_{n-1}(x_0)(s) \right\}^2 dsdt \right)^{1/2} \\
&\quad = K(\varepsilon_{n-1}^2)\|S_{n-1}(x) - S_{n-1}(x_0)\|_{\mathcal{H}},
\end{aligned}$$

by the Cauchy-Schwarz inequality. Repeating the arguments above, we conclude

$$\|S_n(x) - S_n(x_0)\|_{\mathcal{H}} \le \|x - x_0\|_{\mathcal{H}} \prod_{i=0}^{n-1} K(\varepsilon_i^2).$$

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

Taking expectations on both sides and using the independence of the ε_i proves (ii). ■

Theorem 2.2 is a simple corollary to Theorems 2.1 and 5.1. Theorem 2.3 can be proven along the same lines of argumentation, and the proof is omitted.

Proof of Propositions 2.1 and 2.2. First, we establish (2.7). We follow the proof of Theorem 2.1. Since $\mathbb{E}\{\|\sigma_0^2\|_{\mathcal{H}}\}^{\alpha} = \mathbb{E}\{\|Z_\infty\|_{\mathcal{H}}\}^{\alpha}$, according to the construction in the proof of Theorem 2.1 we have

$$\begin{aligned}\mathbb{E}\{\|\sigma_0^2\|_{\mathcal{H}}\}^{\alpha} &= \mathbb{E}\{\|Z_\infty\|_{\mathcal{H}}\}^{\alpha} \\ &\le \mathbb{E}\{\|Z_1(0)\|_{\mathcal{H}} + \|Z_1(0) - Z_\infty(0)\|_{\mathcal{H}}\}^{\alpha} \\ &\le 2^{\alpha}\left\{\mathbb{E}\{\|Z_1(0)\|_{\mathcal{H}}\}^{\alpha} + \mathbb{E}\{\|Z_1(0) - Z_\infty(0)\|_{\mathcal{H}}\}^{\alpha}\right\},\end{aligned}$$

where 0 denotes the "zero function" on $[0,1]$. According the proof of Theorems 2.1 and 5.1, the term $\mathbb{E}\{\|Z_1(0) - Z_\infty(0)\|_{\mathcal{H}}\}^{\alpha} < \infty$. Furthermore, the term $\mathbb{E}\{\|Z_1(0)\|_{\mathcal{H}}\}^{\alpha} = \left(\int \delta^2(t)dt\right)^{\frac{\alpha}{2}} < \infty$. To show (2.10), we note that

$$\begin{aligned}\mathbb{E}\{\|y_0\|_{\mathcal{H}}\}^{\alpha} &= \mathbb{E}\left[\int y_0^2(t)dt\right]^{\alpha/2} \\ &= \mathbb{E}\left[\int \varepsilon_0^2(t)\sigma_0^2(t)dt\right]^{\alpha/2} \\ &\le \mathbb{E}\{\|\varepsilon_0\|_\infty\}^{\alpha}\mathbb{E}\{\|\sigma_0\|_{\mathcal{H}}\}^{\alpha},\end{aligned}$$

since ε_0 and σ_0 are independent processes. Proposition 2.1 is proven.

The proof of Proposition 2.2 only requires minor modifications and is therefore omitted. ■

Proof of Proposition 2.3. Using recursion 2.1, we have

$$\begin{aligned}\omega(y_0,h) &= \sup_{0\le t\le 1-h}\sup_{0\le s\le h}|\varepsilon_0(t+s)\sigma_0(t+s) - \varepsilon_0(t)\sigma_0(t)| \\ &\le \sup_{0\le t\le 1-h}\sup_{0\le s\le h}\big\{|\varepsilon_0(t+s)||\sigma_0(t+s) - \sigma_0(t)| \\ &\qquad + |\sigma_0(t)||\varepsilon_0(t+s) - \varepsilon_0(t)|\big\} \\ &\le \|\varepsilon_0\|_\infty\omega(\sigma_0,h) + \|\sigma_0\|_\infty\omega(\varepsilon_0,h).\end{aligned}$$

The independence of ε_0 and σ_0 yields

$$\mathbb{E}\{\|\sigma_0\|_\infty\omega(\varepsilon_0,h)\}^p = \mathbb{E}\{\|\sigma_0\|_\infty\}^p\mathbb{E}\{\omega(\varepsilon_0,h)\}^p.$$

Proposition 2.2 gives $\mathbb{E}\{\|\sigma_0^2\|_\infty\}^p < \infty$. This implies that $\mathbb{E}\{\|\sigma_0\|_\infty\}^p < \infty$ and therefore

$$\lim_{h\to 0}\mathbb{E}\{\|\sigma_0\|_\infty\omega(\varepsilon_0,h)\}^p = 0.$$

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

The identity $|\sqrt{a}-\sqrt{b}| \le \sqrt{|a-b|}$, $a,b \ge 0$, implies

$$\omega^p(\sigma_0,h) = \sup_{0\le t\le 1-h}\sup_{0\le s\le h}|\sigma_0(t+s)-\sigma_0(t)|^p$$
$$\le \left(\sup_{0\le t\le 1-h}\sup_{0\le s\le h}|\sigma_0^2(t+s)-\sigma_0^2(t)|\right)^{p/2}.$$

Recursion (2.2) gives

$$|\sigma_0^2(t+s)-\sigma_0^2(t)| \le |\delta(t+s)-\delta(t)| + \left|\int(\beta(t+s,r)-\beta(t,r))y_{-1}^2(r)dr\right|$$
$$\le \omega(\delta,h) + \sup_{0\le t\le 1-h}\sup_{0\le s\le h}\sup_{0\le r\le 1}|\beta(t+s,r)-\beta(t,r)|$$
$$\times \int y_{-1}^2(r)dr.$$

Hence,

$$\mathbb{E}\left(\omega^p(\sigma_0,h)\right) = \mathbb{E}\left\{\sup_{0\le t\le 1-h}\sup_{0\le s\le h}|\sigma_0^2(t+s)-\sigma_0^2(t)|\right\}^{p/2}$$
$$\le 2^{p/2}\left\{\left[\sup_{0\le t\le 1-h}\sup_{0\le s\le h}\sup_{0\le r\le 1}|\beta(t+s,r)-\beta(t,r)|\right]^{p/2}\right.$$
$$\left.\times\,\mathbb{E}\left\{\|y_0\|_\infty\right\}^p + \left[\omega(\delta,h)\right]^{p/2}\right\}.$$

Proposition 2.2 yields that $\mathbb{E}\left\{\|y_0\|_\infty\right\}^p < \infty$ and $\mathbb{E}\left\{\|\sigma_0\|_\infty\right\}^p < \infty$. So by the independence of the processes ε_0 and σ_0, we conclude

$$\lim_{h\to 0}\left\{\|\varepsilon_0\|_\infty\omega(\sigma_0,h)\right\}^p = 0,$$

completing the proof of Proposition 2.3. ■

Proof of Theorem 3.1. Under our assumptions, it follows from Theorem 2.2 that for any $m \ge 1$,

$$\mathbb{E}\|Z_k - Z_{km}\|_{\mathcal{H}}^4 \le \text{const} \times r^m$$

where $r \in (0,1)$ and Z_{km} are the m–dependent approximations of Z_k (constructed by using σ_{km}^2 instead of σ_k^2 in the definition of Z_k). This shows that the notion of L^4–m–*approximability* suggested in Hörmann and Kokoszka (2010) applies to the sequence $\{Z_k\}$. As consequence we have with $\hat{c}_i = \text{sign}\langle\hat{e}_i, e_i\rangle$ that

(a) $$\max_{1\le i\le K}\mathbb{E}\|\hat{c}_i\hat{e}_i - e_i\|_{\mathcal{H}}^2 = O\left(N^{-1}\right);$$

(b) $$\max_{1\le i\le K}\mathbb{E}|\hat{\lambda}_i - \lambda_i|^2 = O\left(N^{-1}\right),$$

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

and therefore that

$$\text{(a}') \quad \max_{1\le i\le K} \|\hat{c}_i\hat{e}_i - e_i\|_{\mathcal{H}} = O_P\left(N^{-1/2}\right);$$

$$\text{(b}') \quad \max_{1\le i\le K} |\hat{\lambda}_i - \lambda_i| = O_P\left(N^{-1/2}\right).$$

(See Hörmann and Kokoszka, Thm. 3.2.) The random sign $\hat{c}_i$ (which we cannot observe) accounts for the fact that e_i can only be uniquely identified up to its sign. As our estimator $\hat{\beta}(K)$ doesn't depend on the signs of the $\hat{e}_i$, this poses no problem. We define

$$\sigma_{i,j} = \mathbb{E}\langle Z_1, e_i\rangle\langle Z_2, e_j\rangle$$

and let

$$\hat{\sigma}_{i,j} = \frac{1}{N-1}\sum_{k=1}^{N-1}\langle Z_k, \hat{e}_i\rangle\langle Z_{k+1}, \hat{e}_j\rangle$$

be the empirical counterpart. Then we have

$$\begin{aligned}\mathbb{E}|\sigma_{i,j} - \hat{c}_i\hat{c}_j\hat{\sigma}_{i,j}| &\le \mathbb{E}\left|\frac{1}{N-1}\sum_{k=1}^{N-1}\left(\langle Z_k, e_i\rangle\langle Z_{k+1}, e_j\rangle - \mathbb{E}\langle Z_k, e_i\rangle\langle Z_{k+1}, e_j\rangle\right)\right| \\ &\quad + \mathbb{E}\left|\frac{1}{N-1}\sum_{k=1}^{N-1}\left(\langle Z_k, \hat{c}_i\hat{e}_i\rangle\langle Z_{k+1}, \hat{c}_j\hat{e}_j\rangle - \langle Z_k, e_i\rangle\langle Z_{k+1}, e_j\rangle\right)\right| \\ &=: T_1(i,j;N) + T_2(i,j;N).\end{aligned}$$

The processes $\mathcal{Z}_k = \mathcal{Z}_k(i,j) = \langle Z_k, e_i\rangle\langle Z_{k+1}, e_j\rangle$ are strictly stationary for every choice of i and j, and we can again define the approximations $\mathcal{Z}_{km}$ in the spirit of Section 2. We have by independence of $\mathcal{Z}_0$ and $\mathcal{Z}_{kk}$

$$\begin{aligned}\sum_{h\ge 0}|\mathrm{Cov}(\mathcal{Z}_0, \mathcal{Z}_h)| &\le \mathbb{E}\mathcal{Z}_0^2 + \left(\mathbb{E}\mathcal{Z}_0^2\right)^{1/2} \times \sum_{h\ge 1}\left(\mathbb{E}(\mathcal{Z}_h - \mathcal{Z}_{hh})^2\right)^{1/2} \\ &\le \mathbb{E}\|Z_0\|_{\mathcal{H}}^4 + \mathbb{E}\|Z_0\|_{\mathcal{H}}^2 \times \sum_{h\ge 1}\left(\mathbb{E}(\mathcal{Z}_h - \mathcal{Z}_{hh})^2\right)^{1/2}.\end{aligned}$$

Further, we have by repeated application of the Cauchy-Schwarz inequality that

$$\begin{aligned}&\mathbb{E}(\mathcal{Z}_h - \mathcal{Z}_{hh})^2 \\ &\quad = \mathbb{E}\left[\langle Z_h, e_i\rangle\langle Z_{h+1}, e_j\rangle - \langle Z_{hh}, e_i\rangle\langle Z_{h+1,h}, e_j\rangle\right]^2 \\ &\quad \le 2\left\{\mathbb{E}\left[\langle Z_h - Z_{hh}, e_i\rangle\langle Z_{h+1}, e_j\rangle\right]^2 + \mathbb{E}\left[\langle Z_{hh}, e_i\rangle\langle Z_{h+1} - Z_{h+1,h}, e_j\rangle\right]^2\right\} \\ &\quad \le 2\left\{\mathbb{E}\langle Z_h - Z_{hh}, e_i\rangle^2\mathbb{E}\langle Z_h, e_j\rangle^2 + \mathbb{E}\langle Z_{hh}, e_i\rangle^2\mathbb{E}\langle Z_{h+1} - Z_{h+1,h}, e_j\rangle^2\right\} \\ &\quad \le 2\mathbb{E}\|Z_0\|_{\mathcal{H}}^2\left\{\mathbb{E}\|Z_h - Z_{hh}\|_{\mathcal{H}}^2 + \mathbb{E}\|Z_{h+1} - Z_{h+1,h}\|_{\mathcal{H}}^2\right\} \\ &\quad \le \text{const} \times r^h,\end{aligned}$$

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

for some $r \in (0,1)$. This proves that the autocovariances of the process $\{\mathcal{Z}_k\}$ are absolutely summable. A well-known result in time series analysis thus implies that

$$(N-1)\mathrm{Var}\left(\frac{1}{N-1}\sum_{k=1}^{N-1}\mathcal{Z}_k\right) \le 2\sum_{h\ge 0}|\mathrm{Cov}(\mathcal{Z}_0,\mathcal{Z}_h)| \le c_0, \qquad \forall N \ge 2,$$

(see, e.g., the proof of Theorem 7.1.1. in Brockwell and Davis, 1991), where, as we have shown, the constant c_0 is independent of the choice of i and j in the definition of $\mathcal{Z}_k$. Hence $\max_{1\le i,j\le K} T_1(i,j;N) = O_P\left(N^{-1/2}\right)$.

Using relation (a) above, one can show that also $\max_{1\le i,j\le K} T_2(i,j;N) = O_P\left(N^{-1/2}\right)$. We thus have

(c) $$\max_{1\le i,j\le K}|\sigma_{i,j} - \hat{c}_i\hat{c}_j\hat{\sigma}_{i,j}| = O_P\left(N^{-1/2}\right).$$

We now have the necessary tools to prove Theorem 3.1. By relations (a′), (b′), and (c) we have that

$$\begin{aligned}\left\|\beta(K)-\hat{\beta}(K)\right\|_{\mathcal{S}} &= \left\|\sum_{1\le i,j\le K}\left(\frac{\sigma_{j,i}}{\lambda_j}e_je_i - \frac{\hat{\sigma}_{j,i}}{\hat{\lambda}_j}\hat{e}_j\hat{e}_i\right)\right\|_{\mathcal{S}}\\ &\le \sum_{1\le i,j\le K}\left\{\left|\frac{\sigma_{j,i}}{\lambda_j} - \frac{\hat{c}_j\hat{c}_i\hat{\sigma}_{j,i}}{\hat{\lambda}_j}\right| + \left|\frac{\sigma_{j,i}}{\hat{\lambda}_j}\right|\left\|e_je_i - \hat{c}_j\hat{e}_j\hat{c}_i\hat{e}_i\right\|_{\mathcal{S}}\right\}\\ &\le K^2\left\{O_P\left(N^{-1/2}\right) + O_P(1)\max_{1\le i,j\le K}\left\|e_je_i - \hat{c}_j\hat{e}_j\hat{c}_i\hat{e}_i\right\|_{\mathcal{S}}\right\}.\end{aligned}$$

The proof follows from $\left\|e_je_i - \hat{c}_j\hat{e}_j\hat{c}_i\hat{e}_i\right\|_{\mathcal{S}} \le \|\hat{c}_j\hat{e}_j - e_j\|_{\mathcal{H}} + \|\hat{c}_i\hat{e}_i - e_i\|_{\mathcal{H}}$. ■

REFERENCES

Angelidis, T. & S. Degiannakis (2008) Volatility forecasting: Intra-day versus inter-day models. *Journal of International Financial Markets, Institutions & Money* 18, 449–465.

Aue, A., S. Hörmann, L. Horváth, M. Hŭsková, & J. Steinebach (2012) Sequential testing for the stability of high frequencey portfolio betas. *Econometric Theory* 28, 804–837.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, & N. Shephard (2008) Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.

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

Bollerslev, T. (1990) Modeling the coherence in short-run nominal exchange rates: Multivariate generalized ARCH model. *Review of Economics and Statistics* 74, 498–505.

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

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods*. Springer.

Cyree, K.K., M.D. Griffiths, & D.B. Winters (2004) An empirical examination of intraday volatility in euro-dollar rates. *Quarterly Review of Economics and Finance* 44, 44–57.

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press

Diaconis, P. & D. Freedman (1999) Iterated random functions. *Siam Review* 41, 45–76.
Didericksen, D., P. Kokoszka, & X. Zhang (2010) Empirical Properties of Forecasts with the Functional Autoregressive Model. Technical Report, Utah State University.
Elezović, S. (2009) Functional modelling of volatility in the Swedish limit order book. *Computational Statistics & Data Analysis* 53, 2107–2118.
Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
Evans, K.P. & A.E.H. Speight (2010) Intraday periodicity, calendar and announcement effects in euro exchange rate volatility. *Research in International Business and Finance* 24, 82–101.
Fatum, R. & J. Pedersen (2009) Real-time effects of central bank intervention in the euro market. *Journal of International Economics* 78, 11–20.
Gau, Y-F. (2005) Intraday volatility in the Taipei FX market. *Pacific-Basin Finance Journal* 13, 471–487.
Harrison, J.M., R. Pitbladdo, & S.M. Schaefer (1984) Continuous prices in frictionless markets have infinite variation. *Journal of Business* 57, 353–365.
Hörmann, S. & P. Kokoszka (2010) Weakly dependent functional data. *Annals of Statistics* 38, 1845–1884.
Jacod, J., Y. Li, P.A. Mykland, M. Podolskij, & M. Vetter (2009) Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications* 119, 2249–2276.
Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14, 70–86.
Kargin, V. & A. Onatski (2008) Curve forecasting by functional autoregression. *Journal of Multivariate Analysis* 99, 2508–2526.
Ling, S. & M. McAleer (2002) Necessary and sufficient moment conditions for the GARCH(p,q) and asymmetric power GARCH(p,q) models. *Econometric Theory* 18, 722–729.
Nelson, D.B. (1990) Stationarity and persistence in the GARCH (1, 1) model. *Econometric Theory* 6, 318–334.
Reeder, R. (2011) *Limit Theorems in Functional Data Analysis with Applications.* ProQuest.
Silvennoinen, A. & T. Teräsvirta (2009) Multivariate GARCH models. In T.G. Anderson et al. (eds.), *Handbook of Financial Time Series*, pp. 201–229. Springer-Verlag.
Stout, W.F. (1974) *Almost Sure Convergence.* Acadamic Press.
Wu, W. & X. Shao (2004) Limit theorems for iterated random functions. *Journal of Applied Probability* 41, 425–436.
Yao, F. & H.-G. Müller (2010) Functional quadratic regression. *Biometrika* 97, 49–64.
Zhang, L., P.A. Mykland, & Y. Aït-Sahalia (2005) A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394–1411.

https://doi.org/10.1017/S0266466612000345 Published online by Cambridge University Press